NORTHWESTERN UNIVERSITY

Epistemic Reasons & Cognitive Self-Monitoring

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Department of Philosophy

By

Spencer Paulson

EVANSTON, ILLINOIS

August 2024

Abstract

This dissertation is about the relationship between Epistemology and other domains. In it I want to show that in an important sense, Epistemology is an autonomous domain. That is, epistemic vocabulary (e.g., “knowledge”, “justification”, “rationality”, “epistemic reason”, etc.) cannot be analyzed without remainder into non-epistemic vocabulary. Epistemic phenomena must be explained in terms of epistemic reasons and the form of assessment proprietary to them. Although epistemic vocabulary cannot be reduced to the vocabulary of other domains, Epistemology is nonetheless connected with other domains in an important way. In this dissertation I focus on the connection between Epistemology and the Philosophy of Mind. I distinguish different forms cognitive self-monitoring can take and corresponding forms of cognitive autonomy. The kind of self-monitoring that receives the most attention involves simulating interpersonal argumentation. I argue that the capacity to self-monitor in this way grounds familiar epistemic standards (e.g., knowledge and justification) and explains how reasons get their grip on us. On the picture that emerges, Epistemology and Philosophy of Mind need to be done in tandem. Epistemology gives us the normative framework that is modeled by the self-monitoring system. The self-monitoring system, conversely, explains the kind of activity we engage in that makes that normative framework applicable to us. Epistemic reasons themselves are just episodes of this kind of activity.

In Chapter 1 I argue that epistemic luck cannot be analyzed in non-epistemic terms. Here I generalize Jennifer Lackey’s criticism of Duncan Pritchard’s safety-based account of epistemic luck to all reductive theories. I argue that they will either be extensionally inadequate, or they will illicitly presuppose the vocabulary they are trying to reduce. The upshot is that epistemic luck can only be understood in terms of the epistemic assessment of reasons. By extension, we

can't give a reductive analysis of knowledge. This is not to say we cannot analyze knowledge, just that we can't analyze away the epistemic vocabulary without remainder.

In chapter 2 I explain the relation between epistemic luck, epistemic justification, and knowledge. I argue that previous work in the defeasible reasoning tradition, despite being on the right track, fell short of realizing its full potential. This is because the account of how these three things relate given by its proponents was insufficiently unified, which encouraged many epistemologists to abandon the defeasible reasoning tradition entirely. I argue that a unified and principled account of their relation is possible, and it allows for an elegant analysis of knowledge. In slogan form: good reasons are apparent to the knowing subject.

In chapter 3 I extend the defeasible reasoning tradition to explain how knowledge is gradable. I show that this enables us to bring the debates about fake barns and knowledge from falsehoods to ecumenical resolutions, while also shedding light on a new kind of case. Furthermore, I explain how background assumptions are themselves defeasible and some of the epistemic significance of this fact.

In chapter 4 I make the connection between Epistemology and Philosophy of Mind. I argue that the self-monitoring system of cognitively mature humans uses a model of the normative framework described in the previous chapters to regulate its own cognition. Here I build on that claim to develop an account of the role played by knowledge and justification in doxastic social coordination. I build on the claim that background assumptions are themselves defeasible to develop an account of the foundations of knowledge according to which they are social but not brutely sociological.

In chapter 5 I discuss the relation between the epistemic states of cognitively mature humans, children, and non-human animals. I do this in a way meant to integrate an insight from

virtue epistemology into the reasons-first picture developed here: positive epistemic standing is a state attributable to the subject.

In chapter 6 I discuss the metaphysics of epistemic reasons. I claim that they are episodes of the activity of reasoning. In the preceding chapters, we get a sense of what that activity involves. Here I argue that the materials developed there put us in a position to understand what epistemic reasons are. This account incorporates the insights of rival positions while sidestepping familiar objections to them. The resulting picture is in line with the non-reductive approach to the epistemic developed so far. In fact, it goes further than other "reasons-first" approaches to Epistemology by claiming that reasons are not only basic within Epistemology, but they are also basic full-stop. One key consideration for my argument here is that reasons are, by their very nature, competitors. Although this point is generally acknowledged in the literature on reasons, its consequences for the metaphysics of reasons hasn't been. Participants in the debate about the metaphysics of reasons typically try to identify reasons with something (e.g., facts or mental states) that are either not competitors, or only competitors insofar as they constitute, or bear on the epistemic assessment of, episodes of reasoning.

Acknowledgments

I am very grateful to my committee for their patience as this project took shape. Perhaps the most important thing they did to help me was urge me to be as clear as possible about what I found dissatisfying in extant views in meta-epistemology. Many helpful conversations took place in and around the Epistemological Club at Northwestern. There I also received helpful feedback from other graduate students. Beth Barker, John Beverley, Regina Hurley, Nate Lauffer, Kathryn Pogin, and Jonathan Weid provided particularly helpful feedback regularly without which the dissertation would have been much worse. Among the faculty, Sandy Goldberg, Jennifer Lackey and Baron Reed provided criticism at these meetings that shaped my thinking on the topics considered in the dissertation in important ways.

After many failed attempts to persuade others of the merit of the project, I probably would have abandoned it had it not been for Sandy Goldberg's encouragement and contagious optimism. He must have read over a dozen drafts of the paper that eventually became "Good Reasons are Apparent to the Knowing Subject" on which the second chapter of the dissertation is based. These conversations were perhaps more important for my philosophical development than anything else. A dozen or so other papers of mine related to this project have been improved greatly by his ability to balance charity and criticism. I've found Sandy to be the ideal advisor and I consider myself incredibly fortunate to have been his advisee.

In addition to her many helpful comments on drafts of the papers that became chapters of this dissertation, Jennifer Lackey proved to be an invaluable resource to me in another way. In her review of Duncan Pritchard's *Epistemic Luck,* I saw the sketch of a way to defend a reasons-first position in Epistemology. Working through what she said there helped me find a way to

articulate a point about the priority of reasons in Epistemology that I had been struggling to articulate until that point.

Something similar applies to Baron Reed's paper, "Who Knows?". Working through this paper helped me see the significance of cognitive agency for the theory of knowledge. The fifth chapter of the dissertation has been greatly influenced by that paper. I first became interested in Epistemology after reading Descartes. Talking with Baron reinvigorated my interest in Early Modern Epistemology and helped me see at various points that the ideas I was pursuing were variations on Early Modern themes. This helped me articulate those ideas from new angles, particularly in the fourth chapter of the dissertation and the publication ("Reflective Naturalism") on which it was based.

Peter van Elswyk joined my committee around the time I began to shift my attention to the relation between Cognitive Science and Epistemology. His enthusiasm about the project helped me deal with the doubt I was beginning to experience about the project. Furthermore, his empirically oriented approach to the Philosophy of Mind made it possible for me to engage in a different kind of discussion about my work than I was able to have with the other members of my committee whose questions and concerns tended to focus on the purely epistemological aspect of my project.

Mona Simion joined my committee shortly after Peter. We began talking about my work after I missed the Northwestern/ Notre Dame Epistemology Conference, which I had organized, because I had the flu. In retrospect, I consider myself incredibly fortunate for having gotten the flu right then. Mona quickly became familiar with my work and offered incredibly helpful

comments on an in-progress version of the last chapter of the dissertation. It is difficult to express how grateful I am to Mona for helping me throughout the process of dealing with the job market this past year.

While working on my MA at the University of Houston I received the necessary background in Cognitive Science to write Chapters 4 and 5. For this I am grateful in particular to Cameron Buckner, James Garson, and Josh Weisberg. The paper that eventually became chapter 4 grew out of conversations I had with Cameron Buckner during a seminar on animal cognition. During the last year of my MA Luis Oliveira joined the department and helped orient me in contemporary debates in Epistemology. His subsequent comments on Chapters 4 and 5 (among other papers not in the dissertation) have been incredibly helpful.

I'd like to thank Nate Lauffer for stimulating discussions at least once a week for the last six years. I've spent more time talking to Nate about Epistemology than anyone else, and I'm a much better epistemologist as a result. I still struggle to be as thorough and concise as he is.

Lastly, I'd like to thank *Synthese* for letting me include "Good Reasons are Apparent to the Knowing Subject" and "Reflective Naturalism" as chapters 2 and 4 of this dissertation (respectively) and *Episteme* for letting me include "Luck and Reasons" and "First-Class & Coach-Class Knowledge" as chapters 1 and 3, respectively.

Preface

I began attending the University of Washington just weeks after Laurence BonJour's retirement. Although I never had the chance to take a class with him, his influence was still palpable. One of the first classes I took was an Epistemology course taught by one of his advisees, Lars Enden. I was already familiar with and greatly influenced by Descartes. This course helped connect the cartesian themes with which I was already familiar to contemporary debates about the theory of knowledge and the theory of justification. The course was taught using mostly BonJour's own material and it set the agenda for me going forward.

I was very sympathetic to BonJour's approach. I found his criticisms of externalism compelling, and I found his broadly cartesian orientation congenial. In this course I was first introduced to the Gettier problem, however, and this became a source of irritation for years to come. I wasn't worried so much that it can't be solved as that its solution was bound to make knowledge itself a gerrymandered phenomenon. Initially, I accepted that conclusion reluctantly and came to see Epistemology as primarily the theory of justification. However, I had the lingering sense that there must be something wrong with my account of knowledge if it made knowledge itself the intersection of explanatorily unrelated properties. This concern stuck with me and was one of the two main factors that animated the dissertation. It prompted reflection on the nature of Epistemology itself, whether Epistemology should be done from the perspective of the deliberating agent, and whether epistemic vocabulary could be reduced to non-epistemic vocabulary.

When I went to the University of Houston a few years later to earn my MA, I acquired another concern that contributed in equal measure to the shape of the dissertation. I went to UH

because I had just read Tyler Burge's *Origins of Objectivity.* Inspired by Burge's work, I wanted to learn more about cognitive science and find a way to integrate my views in Epistemology into a broader account of the mind and its place in nature. Most epistemologists with interdisciplinary ambitions felt the need to abandon the broadly cartesian project that I had inherited from BonJour (and, of course, Descartes himself). They thought their interdisciplinary ambitions were best realized by reducing epistemic vocabulary to the vocabulary employed in the sciences, such as information theory, nomic covariation, belief-forming processes and their reliability, etc. I was interested in pursuing a different approach. I was interested in uniting Epistemology and Cognitive Science without reducing either. I had a vague sense that that the broadly cartesian approach had the best chance at vindicating the epistemic standing of cognitive science itself and that cognitive science itself could learn something important from it. But I nonetheless felt that more needed to be said to connect that project with the picture of the mind and its place in nature that was emerging in Cognitive Science than anyone laboring in that camp had yet said. I was inspired by Wilfrid Sellars who seemed to be pursuing something similar and, I still believe, had important insights in the vicinity. However, I don't think he was ever fully satisfied with his work on the matter and cognitive science changed in important ways since that work was done (or, perhaps better, science changed in important ways because cognitive science was born after that work was done).

These two concerns animated the dissertation. To address the first, I felt I needed to clean up the defeasible reasoning tradition itself. I needed to address the worry that its solution to the Gettier problem was *ad hoc* that had been raised by Williamson and others. I needed to show that the justification condition and the anti-Gettier condition are two sides of the same coin, much

like the first- and third-personal perspectives. That points in the direction of a solution to the second problem. By giving us a normative framework for toggling between the first- and third-personal perspectives, the defeasible reasoning tradition provides something that proponents of the "Social Intentionality Hypothesis" in cognitive science have long been presupposing. That is, the defeasible reasoning tradition gives us the normative kinematics of interpersonal argumentation. According to SIH (or one version of it rather), intrapersonal cognitive self-monitoring involves modeling those norms and using them to engage in top-down self-regulation. The project then became building a system around these claims. The dissertation is the beginning of that project.

Table of Contents

Introduction

I started doing Epistemology at a time when the relation between Epistemology and other things was receiving a great deal of attention. Pragmatic encroachment had been around for a while and moral encroachment was coming on the scene. Debates surrounding the above centered on the relation between Epistemology and Philosophy of Action and Ethics (respectively). Epistemologists were also beginning to shift their attention to Social Epistemology. Questions about testimony, peer disagreement, and epistemic social practices more broadly were receiving increased recognition as a core part of Epistemology, rather than afterthoughts that could be addressed after the main theoretical questions had been settled.

Long before I started doing Epistemology, people were investigating the relation between Epistemology and the picture of the mind and its place in nature that was (and still is) emerging from Cognitive Science. Alvin Goldman had, before I was born, tried to usher Epistemology into an interdisciplinary era. By the time I was beginning to do Epistemology writers such as Hilary Kornblith were making a case that epistemic reflection is less philosophically significant than the tradition had taken it to be. If we take the interdisciplinary approach seriously, it is just one cognitive process among others, and not especially reliable, according to some noteworthy experimental results.

The cumulative effect of the above was that I started doing Epistemology at a time of disciplinary upheaval. Philosophy is kind of like perpetual crisis science, so upheaval is to be expected. However, the extent of it seemed greater than was the case in previous decades. There was a general state of dissatisfaction with what was now being called “traditional epistemology”.

That term is a seemingly pejorative one for the project of conceptually analyzing knowledge and epistemic justification as well as, perhaps, thinking of that project as the most philosophically significant task for epistemologists. Traditional Epistemology has been the target of many criticisms, a frequent theme of which is that its disciplinary boundaries are too narrow. To live up to its full potential, Epistemology needs to expand.

Since I came of age during a period of rebellion, my sympathies to the traditional approach will strike many as quaint and old-fashioned. In conversation with Nate Lauffer, I sometimes joke that he and I are the last epistemologists not to begin their epistemological journey suspicious of traditional epistemology. Relatively early in my philosophical development I found myself sympathetic to versions of internalist foundationalism about justification according to which foundationally justified beliefs nonetheless depend negatively on the absence of defeaters. I similarly found myself favorably disposed to defeasible reasoning solutions to the Gettier problem. This latter conviction was not due so much to reflection on counterexamples (although that is sometimes helpful), but the conviction that there is something wrong with the subject's reasons for belief in Gettier cases and knowledge (like justification), is a matter of the quality of the subject's reasons. In Epistemology textbooks from the late 20th century epistemological problems were often framed in terms of problems with our reasons or evidence. Regress skepticism is a problem with a regress of reasons, other kinds of skepticism (external world, other minds, memory, induction, etc.) are problems with the circularity of our reasons for believing certain kinds of things. I find that way of framing epistemological problems compelling and that is part of why I see myself as continuing the project of traditional epistemology. However, the point doesn't depend on excessive concern about skepticism.

Rather, I find those skeptical challenges compelling because they are an extension of (or at least purport to be an extension of) everyday reasoning. It seems pre-theoretically plausible that there is an intimate connection between being justified in one's beliefs and having reasons for them, and there seems to be an important connection between having reasons and giving reasons (although the nature of that connection might be complicated). So, the skeptical challenges draw out a tension in how we pre-theoretically regard ourselves epistemically. The challenge of bringing my commitments into reflective equilibrium drew me into Philosophy in the first place, so the epistemological traditions oriented around skeptical challenges have always appealed to me.

This is a very brief description of what I find attractive in traditional epistemology. I think it would be a shame for Epistemology to become disconnected from the challenges that animate traditional epistemology. However, there is much to be said in favor of a more expansive conception of Epistemology. Sellars said that Philosophy is about how things, in the broadest possible sense, hang together, in the broadest possible sense. I also think that's true, or at least that philosophers ought to make it true. Although the problems of traditional epistemology are worth taking seriously, it would be a shame for Epistemology to treat those problems in isolation of other philosophical problems. The result would be a myopic discipline. Despite the considerable (often meta-philosophical) disagreements among those pushing for a more expansive understanding of Epistemology, there seems to be a shared concern that the potential for myopia was actualized by traditional epistemologists in the late 20th century. I think this charge is largely, but not entirely, fair. One of the goals of this dissertation is to grant the point, even argue for it at some length, but nonetheless extract insights from that tradition that

play a central role in a non-myopic epistemology within which our nature as social animals and the relation between epistemological theorizing and the conception of ourselves that emerges from cognitive science is central to the project of Epistemology itself.

To do this I need to defend traditional epistemology while at the same time reconceiving it. To foreshadow a bit, on the picture that emerges individual epistemology is subordinate to social epistemology in an important sense. Epistemology is social all the way down because reasoning itself is, in the paradigmatic case, interpersonal. So, the analysis of S knows that p is a derivative concern for Epistemology. The analysis of knowledge from the defeasible reasoning is, for the most part, correct. However, the machinery of defeasible reasoning is philosophically significant primarily because of the role it plays in epistemic social coordination. The reflection of the solitary inquirer is a derivative phenomenon that is constituted by simulating reasoning in the interpersonal context. This claim is both empirical and philosophical. I am not deriving an epistemic ought from an evolutionary is, but I am describing (at a certain level of abstraction) a computational process by which a cognizer could implicitly commit herself to familiar epistemic norms and, as she comes to understand the source of those norms, recognize their legitimacy. On the picture that emerges, even the foundations of knowledge are socially inherited (which is not to say that they are a brute sociological fact). Epistemic justification is a winning hand in the solitaire version of a paradigmatically multi-player game. The solitaire version of the game approximates the interpersonal version of the game, and the interpersonal game has the potential to gradually approach its own ideal limiting point at which the distinction between knowledge and justification collapses. The purpose of the game is better achieved as its potential becomes better realized and its players approach the ideal end point of inquiry.

The result is a new understanding of the epistemic significance of reflection, cognitive autonomy/epistemic agency, the relation between inquiry, knowledge and justification, and the relation between epistemology and cognitive science. The result is an expansive conception of Epistemology, but the resources of traditional epistemology are what tie it all together.

The dissertation is the beginning of the project just described. In it I try to put the key pieces in place so that they can be built upon later. Some of that later building has taken place since writing it. I have papers currently under review in which I say more about inquiry and its relation to Epistemology generally, knowledge and its significance for psychological explanation, the form of cognitive autonomy I claim cognitively mature humans enjoy and its implications for strong AI, and the relation between cognitive agency and mental modularity. In writing those papers I came to realize that I have more to say about all of the above topics than I managed to say in those papers, so there will need to be more in the near future. Furthermore, I just recently began to see connections between my account of the internalization of argumentative norms and the internalization of ethical norms. The internalization of rules is an important aspect of ethical evaluation on a number of theories, most notably Rule Consequentialism but also in the Rawlsian theory of justice. The nature of internalization itself tends to be discussed in a frustratingly hand-wavy manner by all parties. In a recent article forthcoming in *Philosophy & Phenomenological Research*, "Internalizing Rules", I advance the discussion by giving an account of what it is to internalize a rule, how I think the ability to do it is (as a matter of contingent fact) implemented in humans, and how we should re-frame a number of theoretical problems as a result.

Chapter 1: Luck and Reasons

Can we analyze the kind of epistemic luck that precludes knowledge in non-epistemic terms? By "non-epistemic," I mean terms that can be understood prior to understanding epistemic evaluation. Safety theories[1], for example, are non-epistemic. The intent of these theories is to explain the kind of luck that interests epistemologists by appeal to a counterfactual analysis that can be understood without first understanding epistemic assessment or the vocabulary proprietary to it (i.e., rationality, evidence, defeat, etc.). Some success-from-ability theories also fit the bill.[2] According to these theories, knowledge-undermining luck arises when a subject forms a true belief by exercising a cognitive ability of the right sort, but the fact that the subject believes truly is not explained (sufficiently or in the right way) by the fact that the agent formed the belief by exercising that ability.

I will start by considering the safety account. I will look at Duncan Pritchard's (2005) version of it and Jennifer Lackey's (2006) criticism of it. The purpose of doing this will be to try to extract a general lesson from the counterexamples she offers and the recipe she uses for generating them. The lesson, I will argue, is that safety theories only give us the correct result if our assessment of the relevant counterfactuals is guided by an explanatorily prior assessment of the subject's reasons. This spells trouble for the reductive ambitions of the safety-theorist. I will then turn to success-from-ability theories and argue that reductive versions are plagued by the same problem. That is, our intuitions about cases are only vindicated insofar as our explanations

[1] See Pritchard (2005; 2007) and Kelp (2013), for example. The worries I raise here will apply just as much to safe-methods views (e.g., Hawthorne 2004) as traditional safety-theories. For this reason, I will use "safety" to refer to a genus of which both are species.

[2] See Sosa (1991; 2007), Greco (2003; 2010; 2012) and Turri (2011). These theorists might not think of themselves as giving an alternative to the safety theory of luck but rather giving an account that entails safety and offers further benefits besides. Even so, if my argument succeeds, they will be shown to fail in these ambitions as well.

of why beliefs turned out true are guided by an explanatorily prior assessment of the subject's reasons for belief. In both cases, the assessment of the subject's reasons will be the kind familiar from defeasibility analyses of knowledge.[3]

The upshot will be that the problems for non-epistemic analyses aren't generated by features that could be Chisholmed away. There is no way to show this conclusively without considering every possible non-epistemic analysis, but a cumulative case will be made that gives us grounds for skepticism about the tenability of the reductive project. This will give some (strong but inconclusive) support for a reasons-first[4] account of epistemic luck.

1.1 Safety

Duncan Pritchard (2005: Chapter 5) offers the following account of luck generally. For an event, E, to be lucky, it must happen in the actual world but not in a wide class of the nearest possible worlds (Pritchard 2005: 128). The other condition on luck is that E must be significant to the agent concerned (Pritchard 2005: 132). The first condition is the one I will focus on here. The basic idea is that lucky events could have easily failed to happen, and we understand ease of failure counterfactually. This is an account of luck generally. It applies to the epistemic and the non-epistemic alike. To explain veritic epistemic luck[5] (i.e., the kind familiar from Gettier cases), he proposes that the truth of a belief is veritically epistemically lucky just in case it is not safe. A belief is safe just in case it is true in the actual world and could not easily have been false. That is, in nearly all (if not all) of the nearest worlds in which the agent forms the belief in

[3] See Klein (1980), Pollock (1986) and Moser (1989). See Shope (1983) for useful overview of the history of defeasible reasoning as a solution to the Gettier (1963) problem.
[4] Here I speak of reasons rather than evidence, though the point could just as well be made in terms of quality of evidence (cf. Klein 2017).
[5] The term is first used in Engel (1992).

the same way, it is true. Pritchard's account of veritic luck is just the result of taking his general account of luck and applying it to epistemology by making true belief the significant outcome.

In her review of Pritchard's book, Jennifer Lackey (2006) offers a series of criticisms as well as a general recipe for generating them. I will proceed by first considering one of her cases, then considering her recipe, and, finally, adding some details to the recipe in order to make clear the exact nature of the problem she has identified. In the next section, I will apply the lessons learned in this section to reductive success-from-ability accounts and show that a similar objection can be made against them as well. This will give us reason to think that the problem lies not just with safety-based accounts, but rather with reductive accounts generally. In learning about what epistemic luck is not, we will shed some light on what it must be along the way.

> *Southernmost Barn*: While entering a Midwestern farming community on her cross-country drive, Janice looked at the first barn that she saw, which was on the southernmost end of the field, and formed the corresponding belief 'There is a barn'. As it happens, the barn she saw is the only real one, surrounded by barn façades that members of this community have placed in the field in order to make their town appear prosperous. However, as a matter of strict and unwavering policy, the members of this community always place their only real barn on the southernmost end of their land since this is where traffic first enters their town. Moreover, thirty years earlier, Janice had lived in a house on the southernmost end of this field in the precise location of the one real barn. Because of her deep interest in her childhood roots combined with the brief period during which she can safely take her eyes off of her driving, she would invariably have looked at only the particular place in the field where the real barn exists. (Lackey 2006)

Lackey claims, rightly I think, that this is a paradigmatic case of epistemic luck.[6] She also points out that the subject's belief is safe. A lot would need to change for her to believe falsely forming a belief in that way. This is the point of introducing the "strict and unwavering policy" of the community and the biographical details about the subject only being disposed to look at a very particular spot. Furthermore, if there were any doubt that the subject's belief is safe, we could simply make the vignette longer and introduce more details to guarantee counterfactual robustness. The key is to make the details sufficient to ground the counterfactual but epistemically irrelevant.

Of course, the counterexample she proposed is not an isolated issue. It is not hard to generate similar ones. Lackey offers us the following recipe for cooking them up,

> Still further, numerous other kinds of counter-examples similar to *Southernmost Barn* can be constructed which fail Pritchard's safety-based view of knowledge. First, choose a paradigmatic Gettier-type case, such as Bertrand Russell's stopped clock example. Secondly, construct the case so that there is a feature *x* such that nearly invariably, when *x* is present, *S* will form the true belief that *p*, and *p* will be true only when *x* is present. Thirdly, ensure that the connection between *x* and *p* is entirely non-epistemic, e.g., *S* believes that *p* only when what is picked out by *p* is pink, and *p* will be true only when what is picked out by *p* is pink. Fourthly, if there are any residual doubts that the belief is

[6] She claims that it is a paradigmatic Gettier case (Ibid). If we wish to distinguish classic Gettier cases from post-Gettier cases, we might quibble with this. Nothing important would hang on the outcome of the quibbling. It is generally agreed that post-Gettier cases (such as Ginet's original fake barn case) are cases of epistemic luck and *Southernmost Barn* resembles Ginet's original case in all the important respects.

> genuinely safe, add further features to guarantee counterfactual robustness across nearby possible worlds (Lackey 2006).

This is correct, as far as it goes. However, we should try to be more precise about what we mean when we say that the connection between two things is epistemic. I will here go beyond what Lackey explicitly says, although my proposal is in the same spirit. To foreshadow a bit, I will draw on the defeasible reasoning literature to unpack the epistemic/non-epistemic distinction. Apart from the arguments that have already been made in favor of centering epistemology around defeasible reasoning[7], we will see that the defeasible reasoning literature earns its keep by enabling us to be more articulate about why some features of *Southernmost Barn* are epistemically relevant while others are not in a way that vindicates the intuitive verdict. Next, we will see that we can use defeasible reasoning to explain why the safety theory gets many cases right, but it gets Lackey-style cases wrong.

I recommend trying to get traction on the epistemic in terms of reasons.[8] Consider *Southernmost Barn* again. The subject has a *prima facie* reason to think that she is looking at a barn: her perceptual experience of it. However, that reason is defeated by the fact that she is in fake barn county. That is, the fact that she is in fake barn county is a *prima facie* reason to give up her belief that she is seeing a barn because it undercuts the relation between her perceptual evidence and the truth of her belief. After all, things would look the same to her even if she were looking at one of the many barn facades in the area. The defeating reason is not part of her

[7] See e.g., Audi (1993), de Almeida & Fett (2016), Klein (2017) and Lehrer (2017).
[8] More specifically, I will do so from within the defeasible reasoning tradition. Some prominent objections to this approach that I will not consider here are those of Feldman (2003), Foley (2012) and Turri (2012). In my view, an ample response to all three can be found in de Almeida & Fett (2016). Lasonen-Aarnio (2010 a, b; 2014) and Baker-Hytch & Benton (2015) raise a distinct set of worries I will not be able to address in this paper.

possessed evidence. So, this reason presents an obstacle to her having knowledge though not to her having justification. This, at any rate, seems plausible. Some varieties of direct realism might say otherwise because they individuate perceptual states in such a way that the perceptual states themselves are different in the good and bad cases.[9] I submit that it would be a bit doctrinaire to deny that perceptual evidence is defeasible (even in the good case). So, I will assume going forward that it is, even if perceptual states themselves are individuated as the direct realist says they are.

It is possible for a subject to have a true belief based on a defeated reason and still have knowledge. Defeaters are *prima facie* reasons to give up a belief. So, they need not be *ultima facie* reasons to do so. Defeaters can themselves be defeated. Sometimes they are defeated in such a way as to restore the epistemic potency of the subject's justifying reason. Other times, the defeater-defeater gives the subject new reasons rather than restoring the old ones.[10] *Southernmost Barn* involves the latter sort. The reason to give up the subject's belief is the undercutting defeater mentioned earlier. However, that reason is trumped by the fact that the subject happens to be looking at the one real barn in the area. However, it isn't clear that this restores the epistemic potency of her perceptual reason. Perceptual reasons involve a general capacity for discrimination. However, what is doing the real work here is not her ability to generally discriminate barns from non-barns but rather the fact that she happens to be looking at the one real barn in the area. Insofar as she is basing her belief on a general capacity for discriminating barns from everything else, she is in trouble. She does not have that general

[9] E.g., McDowell (1994), Brewer (1999). Some direct realists, such as McDowell, want to accommodate the defeasibility of perceptual evidence. See Ginsborg (2006) for discussion.
[10] To my knowledge, Klein (1980) first made this point. See also Pollock (1986: Appendix) and de Almeida & Fett (2016).

capacity in fake barn county. It is true that when she is looking at this one particular real barn, she gets the right answer. Could we say that her perceptual experience is a suitable (for knowledge) reason in rural areas, but not in the narrower reference class of fake barn county, but it is once again suitable in the yet narrower reference unit class with this particular real barn as its sole member?

It is worth noting that perceptual reasons in general only function against certain background assumptions. For our perceptual experiences to count in favor of believing anything, we must assume that lighting conditions are more or less standard, that we haven't been drugged, etc. It may be true that perceptual experiences can function relative to background assumptions much more fine-grained than the general default ones (perhaps even specifiable using demonstratives and applicable only to a single case). Nonetheless, those are not the background assumptions the subject is actually making. In fake barn cases, we are imagining a subject ignorant of local idiosyncrasies. So, we imagine her making the default background assumptions she makes for perception in general. Those default/general background assumptions are defeated by the fact that she is in fake barn county. Nothing about the particular barn she is looking at restores *those* background assumptions to their knowledge-supporting post: they are false. Since the standing of the subject's reasons depends on her (actual) background assumptions, the standing of her reasons is problematic in such a case.

I am in effect granting the earlier proposal that her perceptual experience is a suitable (for knowledge) reason in rural areas generally, but not in the narrower reference class of fake barn county, but it is once again suitable in the yet narrower reference (unit) class of this particular real barn. So, her foreground perceptual reason is undefeated by the facts and consequently

capable of supporting knowledge. However, her background assumptions (to the effect that she is in a normal perceptual environment) are defeated by local idiosyncrasies and never restored later on because they are false. I am not assuming that there can be no knowledge from falsehood (cf. Warfield 2005; Klein 2008). I will talk more about those cases in Chapter 3. I am assuming that fake barn cases are not cases of knowledge from falsehood.

I am taking the familiar apparatus of defeasible reasoning and then applying it separately to foreground reasons and background assumptions. Her foreground reasons are undefeated by the facts because there are other background assumptions relative to which they could function that are not themselves defeated by the facts (these are the demonstratively specifiable ones only applicable to her particular case). Admittedly, she isn't making those background assumptions and isn't in a position to reasonably do so. However, that doesn't impugn her foreground reasons themselves, only her own epistemic standing in relying on them the way she does. The foreground reasons as such are fine so long as there are some undefeated background assumptions relative to which they could function. Nonetheless, the subject can't know on the basis of them because foreground reasons require background assumptions and the assumptions she is actually making are defeated by the facts. I will return to this point in the third chapter to give a more nuanced treatment of this kind of case. There I will have to qualify some of what I say here, but the key points will remain intact after doing so.

It is worth contrasting this with a case where a subject sees a barn in ostensibly normal conditions and believes there is one before her but, unbeknownst to her, the New York Times just published an article saying that the county she is in is fraught with Potemkin barns. The Times report turns out to be false. The testimony from the Times provides an undercutting

factual defeater. It undercuts her foreground reason by rebutting her background assumption that she is in normal perceptual conditions. The fact that the reporter made a mistake is a defeater-defeater that restores her perceptual reason by restoring her background assumptions (conditions are normal after all, the defeater is seen to be a "false alarm" at the end of the vignette). Unlike the traditional fake barn case, the original background assumption that conditions are normal is restored and does not need to be replaced by new, more specific background assumptions. So, unlike the fake barn case, the subject is not a victim of knowledge-undermining luck[11][12].

Let us now return to Lackey's point about the epistemic. We have seen that only facts that bear on the quality of the subject's reasons are properly epistemic. This is what explains our intuitions about Lackey-style cases. When I talk about the quality of the subject's reasons, there are two dimensions of assessment that are relevant. One dimension of assessment has to do with the subject's rationality/internalist justification. For her reasons to do well along this dimension they must be *prima facie* reasons undefeated by her total evidence. However, a subject can lack knowledge despite being perfectly rational. This is because her reasons are defeated by facts of which she is unaware. So, the first dimension of assessment involves surviving the crucible of potential defeat against her total evidence. The second involves surviving the crucible of potential defeat against the facts. When I speak of quality of reasons without specifying either dimension of assessment, I mean to talk about the dimensions jointly. That is, if I say there is an

[11] My own view, which I do not have the space to develop in this paper, is that the subject in a fake barn case has their knowledge somewhat but not entirely undermined by luck, as I will argue in Chapter 3. What matters for my purposes in this paper is that there is *some degree* of epistemic luck in fake barn cases that at least to some degree undermines the subject's knowledge. I succeed so long as I can explain the degree of epistemic luck we find in these cases in terms of how we assess the subject's reasons.

[12] This differs from Harman's (1973) take on this kind of case.

issue with the quality of the subject's reasons, I mean that they are lacking along (at least) one of these dimensions.

We can give an account of what goes wrong in Gettier cases in terms of quality of reasons. What cases like *Southernmost Barn* show us is that a subject's reasons can be defeated (and hence incapable of supporting knowledge) but the belief can nonetheless be safe. To generate cases like this, we just have to weave details into the story that are causally significant but not (restoring) defeater-defeaters. That is, we write the story so that it includes a factual defeater (just like in a Gettier case). Then, we can make the truth of the subject's belief counterfactually robust by adding details to the story that have no bearing whatsoever on the quality of her reasons. The details need to be causally efficacious but not (restoring) defeater-defeaters. The fact that the subject happens to have a nostalgic attachment to one particular region of the county has no bearing on the quality of her reasons. It is a brute psychological fact. It does not have any bearing on the question of whether her background assumptions are capable of supporting knowledge. The fact that the community members invariably put the one real barn in this spot has no bearing on the quality of her reasons either. It is causally significant but not epistemically significant in this case. The subject believes truly and in a counterfactually robust way, but it is still a case of luck.

We can now see why the safety theory has some apparent plausibility as a theory of veritic luck and also diagnose why the plausibility is merely apparent. In the examples Pritchard offers to motivate the theory, the salient features of the vignette are properly epistemic. When he has us consider unsafe beliefs that are not (intuitively) knowledge, the causally salient aspects of the story bear on the quality of the subject's reasons. In classic Gettier cases, for example, the

only salient details are the subject's *prima facie* reasons and the facts that defeat them. Our assessment of the relevant counterfactuals will (ceteris paribus) be guided by our assessment of the subject's reasons. If her reasons are of the sort required for knowledge, then we will be inclined to think the truth of the subject's belief is counterfactually robust as a result. What Lackey has shown is that the epistemic and the counterfactual only align the way Pritchard needs them to ceteris paribus. When we introduce causally efficacious but epistemically irrelevant noise into the vignette, the counterfactuals come out true despite the subject clearly lacking knowledge. That is, when things are not equal, the counterfactual robustness of the belief is divorced from the quality of the subject's reasons.

In fact, we are now in a position to see why a belief can be irrationally held but nonetheless safe.[13] It is because we can introduce causally efficacious but epistemically irrelevant details into the vignette. That is, we can introduce details that play a causal role but don't bear on the quality of the subject's reasons. This is why the safety theory only gives us the right verdict when the causally efficacious factors also bear on the quality of the subject's reasons. Conversely, when the safety theory gets the right result, it is because the counterfactuals are made true by facts about the quality of the subject's reasons in the actual world. In other words, if we don't already understand which facts are epistemically relevant and restrict the causally efficacious factors in our vignettes accordingly, then we will encounter cases such that the subject's belief is intuitively epistemically lucky but nonetheless safe. This shows that the problem is generated by Pritchard's preferred order of explanation rather than a minor detail that could be Chisholmed away. The defeasible reasoning approach scores points not only because it

[13] Thanks to an anonymous referee for pointing out the need to discuss this matter.

is not susceptible to Lackey-style counterexample, but also because it enables us to articulate why the safety theory gets cases right when it gets them right and why it gets them wrong when it gets them wrong, as we have just seen.

The assessment of the subject's reasons must be explanatorily prior to the counterfactuals (insofar as the theory is extensionally adequate) because facts about the quality of the subject's reasons determine the truth of the counterfactuals. If we assume the theory of counterfactuals in Lewis (1973)[14], then the point is that the similarity relation on worlds is explanatorily prior to the truth-values of counterfactuals but explanatorily posterior to the assessment of the subject's reasons.[15] If the quality of her reasons is good in the actual world, then worlds in which she bases her belief on them and believes something false will seem remote to us. Quality of reasons judgments guide modal proximity judgments because the quality of the subject's reasons influences how we see the similarity relation. Other things equal, rationally similar worlds are nearby. If the similarity relation is determined by quality of reasons and the counterfactuals in Pritchard's analysis are determined by the similarity relation, then we understand the counterfactuals (partly) in terms of quality of reasons. This, at least, is what happens in the cases where Pritchard gives us the right answer. However, he cannot (given his reductive ambitions) concede that the assessment of the subject's reasons is prior to the theory of counterfactuals. So, he has no principled way of eliminating alternative similarity relations. Lackey exploits this fact. This is just a restatement of what we have already seen, but in the terminology of Lewis' theory

[14] Which I am doing only for ease of exposition and because Pritchard does as well.
[15] I am not saying that counterfactuals *in general* are explained in terms of reasons, just that the epistemic counterfactuals relevant to the safety theory either will be explained in terms of reasons or the theory will be extensionally inadequate.

of counterfactuals. We have reason to believe it because it is the best explanation of why Pritchard's theory gets many cases right but is susceptible to Lackey-style counterexamples.

We can only state what has gone wrong in Lackey-style cases because we understand the assessment of reasons independently of the counterfactuals that interest Pritchard. It follows that the safety-theorist's order of explanation is what generates the problem. So, it cannot just be Chisholmed away.[16]

There is a worry I should address before proceeding. Those sympathetic to Pritchard will wonder if we can really understand quality of reasons without illicitly relying on a prior understanding of counterfactuals at some point.[17] To address this worry, two hurdles need to be cleared. The account I recommend explains the assessment of reasons by first picking out the *prima facie* good reasons and then using defeasible reasoning to determine whether they are *ultima facie* good reasons. Defeasible reasoning and defeat can easily be shown not to rely on a prior understanding of counterfactuals. The dimensions of assessment involved (i.e., defeat relative to total evidence and defeat relative to the facts) both have to do with rational relations between things in the actual world. The relation between defeater and defeated is a rational relation, so the view does not reduce reasons to something non-epistemic. It also does not rely on counterfactuals since it is actual world defeat that is of interest here and, as I have been arguing, the epistemic counterfactuals that interest Pritchard are explained in terms of rational facts about the actual world rather than vice versa.

---

[16] Lackey (2006) says things to this effect more than once, although her discussion of the point is brief and her diagnosis less committal than mine.

[17] Thanks to Sandy Goldberg for raising this concern.

However, we could still worry about *prima facie* reasons. In virtue of what are some reasons *prima facie* good and others not? *Prima facie* reasons are epistemic foundations. So, to answer the question I will need to give an adequate foundationalist account. I need to do this eventually (see Chapter 4), but I cannot do so here. I can mitigate worry some though by pointing out the following. If some version of internalist foundationalism is tenable, then this worry can be addressed. Internalist foundationalisms do not explain the epistemic potency of the foundations in non-epistemic terms. *A fortiori*, they do not give a reductive counterfactual account of the foundations. So, my conclusions in this paper are conditional on the tenability of some version of internalist foundationalism. It is worth noting in this connection that Pritchard himself is an internalist about justification, he just denies that knowledge entails justification (see his 2005). So, he too is committed to the claims I am assuming here.

1.2 Success-From-Ability

When Stephen Curry makes a three-pointer, the shot is made because of his skill. Curry's skill explains the good outcome. When I make a three-pointer, it is quite often not because of my skill. I make a lot of mistakes and sometimes they happen to off-set each other in just the right way. This suggests a general schema. There are skills (i.e., abilities, competences, virtues) and there are successful outcomes that the exercise of the skill is supposed to bring about. Archery is a skill, hitting the bullseye is a successful outcome. Shooting baskets is a skill, making a basket is a successful outcome. Sometimes the exercise of the skill explains why the successful outcome is obtained. Sometimes the skill is exercised, and the successful outcome is obtained, but the latter is not explained by the former.

A number of authors[18] have tried to make sense of epistemic luck in terms of this schema. They argue in one way or another that a subject knows just in case her true belief is formed by exercising an ability or competence and the fact that she exercised that ability explains why she ended up with a true belief. It is not lucky that the subject's belief is true when she knows. If this schema can shed light on the absence of luck, then it can also shed light on the presence of luck. If the subject's ability does not explain why her belief is true, then the truth of her belief is lucky. The belief is consequently not knowledge.

The basic idea is clear enough. Consider the Nogot/Havit case from Lehrer (1965). The Professor formed the belief that Nogot owns a Ferrari because she saw him driving one (i.e., she exercised her perceptual competence) and he told her he owned it. She existentially generalized and came to believe that someone in her class owns a Ferrari (exercising a competence of deductive reasoning). The resulting belief was true, but only coincidentally. Mr. Havit, another student of hers, happened to really own a Ferrari. Nogot was lying. He was borrowing his rich uncle's Ferrari. So, the professor formed a true belief by exercising her competences, but the explanation of why she believed something true is peripheral to all this. She believed something true because another student of hers happened to own a Ferrari.

The basic idea seems right. The question is whether there is an account of explanation in the offing that can vindicate it without tacitly relying on a prior understanding of quality of reasons. The kind of success-from-ability account I am considering here is supposed to be reductive. It is supposed to explain epistemic luck in non-epistemic terms.[19] This only works if

[18] Among them Sosa (1991; 2007), Greco (2003; 2010; 2012), Turri (2011).
[19] So, knowledge-first virtue epistemologists (e.g., Kelp 2013; Miracchi 2015) are not the target here.

we can understand the relevant sense of the non-epistemic terms without a prior understanding of any epistemic terms. I claim that if we really do understand luck in non-epistemic terms, we will get the wrong verdicts on cases. If we get the right verdicts on cases, we will end up illicitly presupposing the vocabulary of epistemic assessment that is supposed to be reduced. In this way, this section mirrors the previous one.

To see if there really is such a dilemma, let us consider some prominent versions of the theory. Ernest Sosa (2007) tells us that the subject knows just in case her belief is accurate (i.e., true) because it is adroit. Adroit beliefs issue from competences (basically skills or abilities). How does the adroitness need to figure into the explanation of the accuracy? Sosa takes a permissive stance here. The adroitness of the belief needs to be an important part of the explanation of its truth. However, it does not need to be the most important. It can just be one of many important factors.

Sosa makes a distinction between a competence explaining why a subject has a belief and a competence explaining why that belief is true to get around the classic Gettier cases. In Lehrer's case from earlier, Sosa's account maintains that the professor's competence only explains why she formed the belief but goes no distance toward explaining why she achieved accuracy in forming it. This is true, but it does not help in post-Gettier cases where the competence explains not just why the subject has a belief, but also why it is an accurate one. In barn façade county, the subject sees a real barn in normal lighting conditions, etc. Her perceptual competence is an important part of the explanation of why it is a true belief: she saw it in standard lighting conditions where things appear as they are. Sosa is aware of this consequence and comfortable biting the bullet (2007: 96 fn. 1). This plays into the hands of my dilemma. I

claim that the reductive theorist can either illicitly rely on defeasible reasoning or they will be unable to vindicate intuitive verdicts about cases. Sosa took the second horn. Furthermore, Sosa's permissive stance allows for certain self-fulfilling prophecies to count as knowledge when, intuitively, they are not. Suppose I am raising a child and teaching them to perform inductive inferences. Suppose the first time they do so competently, I want to reinforce the behavior by rewarding it. Unfortunately, the conclusion they draw is false (despite being well-supported). So, I change the world to make the conclusion true and then congratulate the child for getting it right.[20] The child's competence is an important part of why they believed something true. The fact that the child believes truly is nonetheless epistemically lucky.

John Turri (2011) also takes this horn (regarding fake barns, not self-fulfilling prophecies).[21] He claims that in cases of knowledge, the truth of the subject's belief is a manifestation of her cognitive virtue. In Gettier cases, the subject's true belief is merely caused by the virtue, but not a manifestation of it. He claims that we should rely on our "robust pre-theoretical" (Ibid) understanding of manifestation to understand the difference. He is right that we have a solid, pre-theoretical understanding of manifestation vs. mere causation generally. The fact that a safe is fire-resistant might cause me to buy it. Nonetheless, my buying the safe is not a manifestation of its disposition to endure fire. When the building burns and the safe is still standing, that is a manifestation of its disposition to endure fire.

[20] This is similar to the "Guardian Angel" case from Greco (2012).
[21] Which is not to say his view is the same as Sosa's, only that they both take the non-standard position that the subject knows in fake barn cases. Turri offers independent argument for reconsidering our intuitions, so it might not be exactly fair to say that he is biting bullets here. Nonetheless, the conclusion is tough for most people to accept.

We should worry that insofar as we get the intuitively correct results, we are relying on more than just our pre-theoretical understanding of manifestation generally. In the classic Gettier cases, we are inclined to say that the truth of the subject's belief is not a manifestation of the subject's competence, but merely caused by it. Why? Turri says it is just our pre-theoretical understanding of the manifestation relation, which is not further analyzable. However, we notice a general pattern. When there is an issue with the quality of the subject's reasons, we will say that we are dealing with mere causation. When there are no concerns about the quality of reasons, we will say that we are dealing with a case of manifestation. This gives the impression that we partition cases based on our understanding of a distinctively epistemic kind of assessment and it guides our judgments about manifestation when we are dealing with epistemic matters. Our prior understanding of reasons makes us a great deal more articulate about why we intuit manifestation where we do. However, our understanding of manifestation in general does not make us any more articulate about quality of reasons. If we could ground epistemic assessment in a prior understanding of manifestation, then we should be able to *derive* facts about quality of reasons from facts about manifestation. However, we cannot, according to Turri, say much about the nature of manifestation and why it shows up where it does. We just rely on our intuitive understanding of it to partition cases. So, the prospects for this look bleak. In fact, if Turri were to derive facts about manifestation from something more fundamental, he would be giving up on his preferred order of explanation. Recall that according to him the manifestation relation is metaphysically and explanatorily primitive.

However, if we go in the opposite direction, a different picture emerges. We can derive facts about manifestation from facts about quality of reasons. This is just to say that if the

subject's reasons do well along the two dimensions mentioned earlier, then the truth of her belief is intuitively attributable to her cognitive abilities and not just caused by them. In Turri's preferred terminology, the subject truly believing is a "manifestation" of those abilities. I submit that the best explanation of the above is that Turri's order of explanation is the reverse of what it should be. The defeasible reasoning tradition gives us the resources to explain why the subject believing truly is differently attributable to her cognitive capacities in the two cases: they afford her reasons of differing quality. The defeasible reasoning tradition can go a step further using the apparatus of defeat to explain why they differ in quality. Turri's proposal can't offer anything comparable.

The upshot is that Turri gives us the worst of both worlds. In fake barn cases where he gives us a principled answer that really does only depend on our understanding of manifestation generally, we get an intuitively incorrect verdict. When we get the intuitively correct verdict in classic-Gettier cases, the verdict is grounded in an independent understanding of the quality of the subject's reasons.[22]

John Greco (2003; 2010) offers us a similar picture. Greco holds that a subject knows p just in case that subject believes the truth (with respect to p) because she formed a belief that p by exercising an intellectual virtue (Greco 2010: 71). An intellectual virtue here is understood in the same way as abilities/competences. Greco says that, in general, a full explanation of an event requires many causal factors. However, some of those factors are going to be more salient to us (the explainers) than others. This is a function of their relative abnormality and our practical

[22] Turri has more recent work on the topic I cannot go into here. See Turri (2016) and Carter (2014) for what I take to be a good response.

interests (2003; 2010). The presence of sparks and the presence of oxygen are both causal factors contributing to a warehouse fire. Nonetheless, the sparks are more salient than the oxygen. The presence of oxygen is typical. It is there on the days when there are no fires. The sparks are a different story. Sparks are unusual. So, they are more salient in our explanation. Relatedly, it is difficult for us to control the presence of oxygen but much easier for us to control the presence of sparks. So, we have a practical interest in focusing on the sparks rather than the oxygen in our explanation.

Greco thinks he can use the above to shed light on Gettier cases. Given our practical interests in exchanging information, it makes sense that intellectual virtues have a default salience for us. So, in a normal case of learning through perception, the subject's perceptual ability will be the most salient part of the explanation of why she believed truly as a result. In Gettier cases, on the other hand, some abnormal events are even more salient. For instance, the fact that Nogot was lying and that, coincidentally, Havit owns a Ferrari are more salient due to their abnormality. Since those are the most salient parts of the explanation of why the professor believed truly, her intellectual virtues have been demoted. So, she does not know. Even though her virtues are a part of the story, they are not a sufficiently salient part of it.

Later (2012), Greco recognized that some kinds of abnormality matter and others do not. He offers the following example to make this point. Suppose I hear my child shriek and turn around to see what happened. I then see that the cat has a bird in its mouth. The sound of the shriek is abnormal. It doesn't seem that this abnormality has any bearing on whether I know that the cat has a bird in its mouth, however. Something abnormal got my attention, so in that respect

I might be lucky to have acquired the evidence I in fact did through perception. Nonetheless, there is no knowledge-undermining luck here.[23]

Greco does not think that this is the most pressing problem for his view, so he puts it aside after mentioning it briefly. I disagree. I urge that this problem is the same one we keep seeing. If it is, it cannot be solved by Chisholming because it stems from the reductive ambitions of his theory rather than the details.

I can best make this point by making a comparison with the earlier discussion of Pritchard. The problem for Pritchard is that the truth of the subject's belief can be as counterfactually robust as we like (relative to its actual basis) and still be the result of knowledge-undermining luck so long as the counterfactual robustness is the result of factors that have no bearing on the quality of the subject's reasons. Someone could form a true belief about the number of a winning lottery ticket by just wishfully thinking about her favorite numbers. With enough ingenuity, we can make the fact that these are her favorite numbers as counterfactually robust as we want. We can make the fact that those numbers were on the winning ticket for this drawing as counterfactually robust as we want (e.g., by telling a story about the number generator's algorithm and the previous states of the machine). We can tell a story about why she would only have formed that belief about the winning ticket for this drawing by making that particular date significant to her in a way that could not have been different unless the whole course of her life had also been different. None of this changes the fact that she made a lucky guess. And we can change the example a bit and make it a Gettier case. The key

[23] It is of the same general sort as Nozick's (1981) "Jesse James" case. Cf. Engel's (1992) novelist case.

point is that we can make the truth of her belief counterfactually robust *in the wrong way* for knowledge. We can make the truth of the belief modally secure and lucky, so long as we off-set the defeaters with causally compensating but epistemically irrelevant factors.

The same problem arises in a different form for Greco. Abnormality, as such, does not allow us to distinguish between epistemically relevant abnormality and epistemically irrelevant abnormality. No amount of Chisholming will change this. The reason is that Greco wants a reductive account of epistemic luck. He wants to explain it in non-epistemic terms. This forces him to traffic in terms like "abnormality." Abnormality, as such, is not a distinctly epistemic property. So, it suits his reductive purposes. However, for the same reason, it does not enable us to distinguish between the kind of abnormality that matters for epistemology and the kinds that do not. It seems clear enough why the abnormal shriek did not undermine my knowledge: it did not amount to a defeater! However, he can't say that. Abnormalities as such are too blunt an instrument to help us distinguish between cases when you are lucky to have the evidence you do and cases where it is a matter of luck that your belief (formed for the reasons it was) is true. Greco cannot pick out the right class of abnormalities without shirking his reductive ambitions.

Greco is right that some events have a default salience in our explanations of why a subject believes truly and that this salience can be overridden by abnormal factors. He is right that this is a general fact about explanation and in no way specific to epistemology. However, he is wrong to think that this gives him everything he needs for a reductive account of epistemic luck. The reason is that in order to understand the relevant sense (for epistemology) of *default salience* and *abnormality*, you must first understand defeasible reasoning. Some reasons give us *prima facie* justification for a belief. If we end up with a true belief based on them, they have a

default salience in the explanation of why the subject believed something true. Nonetheless, the justification (and consequently the salience) can be overridden by a defeater. The default salience of her *prima facie* reason is trumped by the greater salience of the defeater.

In some sense, defeaters are abnormal. That is why *prima facie* reasons are good reasons, at first glance. They do the trick unless something unusual happens. However, to specify the kind of unusual event that matters, you must have recourse to the vocabulary of defeat. This is because we are interested in a specifically epistemic kind of abnormality. Defeaters are only salient to us insofar as we are interested in epistemology. Insofar as we are interested in epistemology, other kinds of abnormality don't matter at all.

It is also worth noting that defeaters are *prima facie* (not *ultima facie*) reasons to give up a belief. So, something can be abnormal in the epistemically relevant sense of "abnormal*"* without undermining knowledge. This happens when the defeater is itself defeated in such a way as to restore the subject's original justification (cf. Lehrer & Paxson's, 1968 "Tom Grabit" case). So, for an account like Greco's to work, we would not just need to find a way to delineate the relevant sense of "abnormal" without using any epistemic vocabulary. We would also have to explain why certain iterated abnormalities undermine knowledge and others do not. We would have to do this without mentioning the rational relations between the abnormalities. Yet the rational relations are the only ones that matter.[24]

[24] This also puts pressure on related accounts that depart from Greco's by maintaining that success from ability is a matter of degree, such as that of Carter (2014). The problem isn't that Greco ignores gradations of success from ability. Rather, it is the order of explanation to which he is committed.

Greco (2012) has offered us a revised account. He now says that the truth of a subject's belief is attributable to her abilities just in case her abilities contribute to believing the truth in a way that could regularly serve the interests of the community. The problem with Gettier cases is that crucial aspects of the explanation are too idiosyncratic to be consistently exploitable.

I see no reason why we can't generate recurring Lackey-cases. Suppose I (for complicated biographical reasons) suffer a compulsion to look at the same clock every day at noon. My circadian rhythm is acute, sub-personal processes cause me to look at exactly noon even though I have not yet formed the belief it is noon. The clock has long been stopped at 12 and (for complicated historical reasons of which I am ignorant) this could not change without much else also changing. When I base my belief that it is noon on what I see looking at the clock, I am Gettiered. I am lucky to get the time right, but I do not know the time. Nonetheless, I'm always right and after forming the belief that it's noon, I walk through town and tell everyone the time. If they could set their watches to Kant's walks, they could set their watches to mine. The classic Gettier cases have a one-off character. However, that is an accidental property. Perhaps the townspeople could not come to know the time on the basis of my testimony. They could consistently get the right answer, despite being lucky to do so. Their reasons would be defeated. My testimony is undercut by the fact that the clock I based my belief on was stopped. None of the biographical idiosyncrasies need restore my reasons for belief to ensure that I always check the clock at the same time. The possibility of regularly exploiting a source of information only requires that certain regularities be in place between the state of the source and the state of whatever the information is about. Those regularities may or may not have any bearing on the quality of the subject's reasons.

1.3 Conclusion

I began this chapter by drawing out the implications of Lackey's criticism of Pritchard's (2005) safety-based account of veritic epistemic luck. I then showed that the problem she identified is an instance of a more general problem that plagues other reductive accounts of epistemic luck. The general lesson is that reductive accounts of epistemic luck only deliver the intuitively correct verdicts about cases when they tacitly (and illicitly) rely on a prior understanding of the quality of the subject's reasons. I showed how this allows us to diagnose the shortcomings of a number of well-known accounts in the success-from-ability tradition.

The assessment of the subject's reasons takes place along two dimensions, relative to the subject's total evidence and relative to the facts. The purpose of the next chapter is to explain the relation between these dimensions.

Chapter 2: Good Reasons are Apparent to the Knowing Subject

I'm interested in making sense of the following two claims:

1) The subject's reasons rationalize her belief.
2) Those same reasons, when all goes well, turn her true belief into knowledge.

These claims both seem very plausible, yet it is difficult to adequately account for them jointly. The problem is that (1) seems to require that we understand reasons in such a way that it is difficult to make sense of (2). If we understand reasons[25] in such a way that their essential property is to rationalize belief, then it seemingly must be an accidental property of theirs that turns true belief into knowledge. This is because Gettier (1963) has shown us that rationality doesn't turn true belief into knowledge. Our interest in knowledge is an interest in the subject's reasons, though it isn't just an interest in whether they are good reasons. It is an interest in whether they are good reasons that stand in the right relation to the facts: something ostensibly foreign to their role of rationalizing belief. To be clear, I am not saying that theorists intend for these roles to be foreign to each other. Rather, I am saying (and will argue below) that we haven't yet seen an account that vindicates our sense that they are not foreign to each other.

This is unfortunate. Knowledge seems too much like the intersection of unrelated concerns to be a philosophically interesting phenomenon. At the very least, it would be better to have an account of how these ostensibly unrelated aspects of the subject's reasons are really distinct aspects of a single concern. Otherwise, the resulting account is in danger of sounding *ad hoc* (cf. Williamson 2000: 31, Pollock 1986: 183, Kvanvig 2003, Schroeder 2015). It would lack

[25] Before getting started I should mention that I like to put my point in terms of reasons, though for my purposes it would also be ok to put the points I make in terms of evidence.

the elegance and theoretical illumination we should hope for (cf. Zagzebski 2017). If knowledge is merely the intersection of otherwise unrelated properties, it is unclear why we care so much about that intersection (cf. Williamson 2000: 31).

I will argue that distinguishing subjective and objective senses of "good reason" to tackle (1) and (2) (respectively) and requiring that the knower's reasons be both is not enough. Very briefly, the subjective sense of "good reason" or "normative reason" has to do with rationality and the objective sense has to do with some kind of cosmic serendipity. These seem unrelated to one another and simply requiring that they coincide doesn't address that problem. Why is it those two things, rather than two other things, that must coincide? I will argue that a similar worry arises for attempts in the defeasible reasoning literature according to which (1) is a matter of justification and (2) is a matter of that justification being "robust" (Schroeder 2015, "defeat-proof" (de Almeida & Fett 2016, Klein 2017) or "truth-resistant" (Moser 1988, 1989).

The key is to understand the "cosmic serendipity" aspect of knowledge in such a way that it is distinct from but not unrelated to the role that reasons play in rationalizing the subject's belief. This will give us an illuminating account of how the two aspects of the subject's reasons that interest us hang together. I will end by giving an account that does this by showing how the sense of "good reason" required to make sense of (1) is related to the sense required to make sense of (2). More specifically, a reason that is good in the former sense is apparently good in the second. Once this idea is unpacked, we will see that our interest in knowledge is an interest in whether the subject's reasons really are as they appear. Good reasons (in a sense to be articulated) are apparent (in a sense to be articulated) to the knowing subject.

In section (I) I will briefly discuss claims (1) and (2). In section (II) I will discuss how our efforts to make sense of (1) seem to put us in a difficult position with respect to (2). In section (III) I will critically assess what I take to be the best attempts to address this difficulty. In section (IV) I will present my own account of how our interests in (1) and (2) relate and the picture of knowledge that emerges.

2.1 Reasons in Epistemology

The first of the two claims that I want to consider is that the subject's reasons rationalize her belief. By "rationalize" people sometimes mean a *post hoc* confabulation. This is not what I have in mind. Rather, I mean that the subject's reasons make it the case that her belief is rational, at least in the case in which her belief is rational. Basic or foundational beliefs present a complication here. I will discuss that in section (IV).[26] In any event, what draws many of us into epistemology is that we wonder whether we are as rational as we take ourselves to be (cf. Fumerton 1995, Audi 1993, BonJour 2003: 40, 2010, Smithies 2015). We want to know if our beliefs are rational and if the question of rationality doesn't arise for some beliefs, then we want to know what is special about them.

When we are looking into the rationality of a subject's belief, we look at how reasonable that belief is. If her reasons make her belief reasonable it is because of the state of her total evidence. I am having a perceptual experience as of a Kentia Palm. Is that a good reason for me to believe that there is one where I am looking? Relative to the total evidence I in fact have, it is. However, if it were part of my evidence that I am somewhere densely populated with artificial

[26] Non-foundationalists say that reasonableness always requires having further reasons. Foundationalists deny this. I will remain neutral until later in the paper.

Kentia replicas, then it wouldn't be such a good reason. Put another way, relative to the evidence I in fact possess, I have sufficient subjective normative reason to believe that there is a Kentia Palm where I am looking. If it were part of my evidence that there are many artificial replicas nearby, then I would not have sufficient subjective normative reason to believe it. "Subjective" here simply means that the normative reason is sufficient insofar as it is being assessed relative to the subject's total evidence. [27]I use "good reason" and "sufficient reason" interchangeably in this paper. I do this for stylistic reasons. At certain points "sufficient" would be clunky while "good" sounds fine.

The above is a very natural line of thought and there is surely something right about it. The basic idea is that we start with the idea that reasons rationalize beliefs. Our recognition that they play this role informs our understanding of what a good reason is. I take this idea about the connection between reasons and rationality to be a well-ensconced part of common-sense that is worth trying to preserve.

Epistemology is also about knowledge. How does this relate to our interest in rationalization? At first glance, there seems to be a very close relation. There are a few ways to see this point. When we ask someone how they know something, we are asking for their reasons (Pollock 1986: 5, Scanlon 1998). This suggests (2). If we answer that question by giving our reasons, this suggests that when our true beliefs are knowledge, it is because of our reasons. A related point comes up when we try to explain why more than mere true belief is needed for knowledge. Someone could acquire a true belief by making a lucky guess. Surely this isn't

[27] As opposed to the facts, in which case we would be considering its sufficiency as an objective normative reason. More on objective normative reasons and their relation to subjective normative reasons below.

knowledge. What is missing? The obvious answer is that they didn't have any good reason for believing as they did (cf. Steup 1996: Chapter 1, BonJour 2010: Chapter 2).

If this much were right, then we would have an adequate account of (1) and (2). In order to make sense of (1) we come to believe that a good reason is one that rationalizes the subject's belief. A good reason, so understood, is also what turns a true belief into knowledge. Our interest in knowledge over and above true belief is an interest in whether the subject has good reasons. It seems hardly surprising that we would be interested in knowledge, if that is what the interest comes to. Offering and assessing reasons is an important aspect, perhaps the most important aspect, of our intellectual lives.[28] Our capacity for this kind of assessment makes us persons.

What I am doing here is taking our practices at face value and then trying to see if we can vindicate them (cf. BonJour 2003: 40 for a similar view about the task of epistemology). When we ask how someone knows, we ask for her reasons. The assessment of reasons is second nature to us. We understand this kind of assessment in all its nuance, at least implicitly, very well. If we can understand both (1) and (2) in terms of a single kind of assessment that is second nature to us, then we have shed light on (1) and (2) in such a way as to vindicate our practice at least in the following way. Our way of inquiring into whether someone knows presupposes that her knowledge is a function of her reasons. Our appreciation of reasons, our rationality, is a deep feature of how we understand ourselves. If we understand (2) in terms of the role reasons play in making (1) true, then we understand our interest in knowledge in terms of an aspect of ourselves that we take to be very important to who we are. The centrality of knowledge to our social

[28] For various takes on this idea see Sellars 1956/1997, BonJour 1985: 7, Korsgaard 1996: Lecture 3, Scanlon 1998: Introduction, Moran 2001: Chapter 2, Boyle 2012, Smithies 2016). For a dissenting view, see (Kornblith 2012).

commerce and private inquiries is something that makes sense to us and that we endorse once we see what knowledge is. Once we understand it this way, we see that it makes sense that we care so much about it. This all pulls in the direction of a "reasons-first" epistemology (see Schroeder 2015, 2021; Kvanvig 2017). We get what we want by making sense of knowledge in terms of our prior understanding of reasons and their assessment.[29]

2.2 Hybrid Knowledge

The problem with the picture that emerged at the end of the last section is that good reasons, so understood, are not what turn true belief into knowledge. This is the upshot of the Gettier (1963) problem. The subject doesn't just need good reasons, the world needs to make her belief true in a way that lines up with those reasons. This is all purposely vague, since it is an attempt to state the problem without presupposing any particular solution to it. After setting up the problem in its full generality, I will argue in the next section that the most promising solution on the market comes up short.

Our interest in knowledge over and above true belief still seems to be an interest in the subject's reasons. The Gettier problem is a problem with the subject's reasons not lining up with the facts in the right way. I will continue to assume that rationality is required for knowledge. So, for the subject to know, she must have good reasons (i.e., ones that rationalize her belief). In addition to being good, those reasons need to harmonize with the facts in the right way.

So, our interest in knowledge is not just an interest in whether the subject's reasons are good. It is an interest in whether they are good reasons with some further property that is

[29] Some readers may worry that this is inconsistent with out tendency to attribute knowledge to small children and animals. I disagree, although the point will have to wait until Chapter 5.

orthogonal to their goodness as reasons. In other words, it is not merely an interest in their essential properties but their accidental ones as well. This point does not depend on the ontology of reasons. When I talk about their essential properties here, I simply mean how they perform with respect to the role they are supposed to play as reasons.

A difficulty emerges here. Why is there a single phenomenon with both essential and accidental properties of reasons as components? Put another way, why do the goodness of the subject's reasons and this other property of theirs compose a single, unified phenomenon: knowledge? If this other property of reasons is another sense of "good reason", the problem rears its head in a slightly different form: we have two lexical entries for "good reason", how do they relate? Without an answer, I worry that the concept of knowledge appears to be gerrymandered. An object is gerrymandered just in case it is the mere sum of distinct objects. The objects countenanced by mereological universalists are like this. My hiking shoes and the Queen's eldest Corgi have no interesting relationship. There is nothing in virtue of which they belong together. So, the "object" that is their sum is gerrymandered. Something similar can apply to composite concepts. If their components aren't interestingly related, they are gerrymandered.

My worry is that a certain way of distinguishing the components of knowledge makes it appear to be one of these concepts. That way of distinguishing them seems to be inevitable. Rational subjects have good reasons. Otherwise, they wouldn't be rational. Yet this leaves open the possibility that their good reasons won't line up with the facts in the right way for knowledge. The problem is that it is hard to see how this further requirement of "lining up" relates to rationality.

The interest in rationality seems to be self-contained. It doesn't seem to be a component part of a larger interest. It is a matter of how the evidence within the subject's perspective functions and doesn't seem to bear any interesting relationship to how those reasons function relative to facts beyond the subject's ken. The tendency to give a theory of justification/rationality that abstracts away from how the subject's reasons line up with the facts makes it seem, perhaps contrary to the intentions of the theorists doing this, that we can understand each without reference to the other (more on this below). So, although the anti-Gettier condition is supposed to be the connection between justification/rationality and truth[30], the connection is starting to look like a relation between largely independent concerns (i.e., what the subject's evidence makes it reasonable for her to believe, on the one hand, and how many important worldly facts she happens to be aware of on the other).

We should try to state what it is that concerns us when we inquire into knowledge in such a way that we can then see both aspects of the subject's reasons as distinct but complementary aspects of it. Saying they are interestingly related because they are components of knowledge presupposes what we would like to have vindicated: that knowledge is a unified, philosophically interesting phenomenon. To be clear, it isn't surprising that we sometimes find gerrymandered properties practically interesting. For practical purposes it sometimes makes sense to fixate on a cluster of essential and accidental properties (e.g., the previous owner of a baseball bat). But it

[30] Perhaps better, a (rather than the) connection between justification and truth. Justification is already connected to truth in the minimal sense that your justification is something that makes it rational for you to regard something as true. The problem is that your justification needs another connection to truth for you to have knowledge; this is the upshot of the Gettier problem. Another way of voicing the concern is that we have two different truth-connections, and they seem to both be interesting but not like two different aspects of a single interesting thing.

would be strange to find such a motley collection philosophically interesting enough center an entire discipline on the study of it.

It is possible that we will never be able to vindicate our sense that the components of knowledge are interestingly related. Perhaps our inquiry into knowledge will embarrass us just as our inquiry into Jade did.[31] If this is where we end up, then we know which properties intersect to form knowledge, but it becomes mysterious why we take such philosophical interest in that intersection (cf. Williamson 2000: 31 for a related point[32]). If our account is a mere aggregate of ingredients, then it lacks depth (even if the aggregated properties are individually deep).[33] The list of ingredients seems *ad hoc* until we give a principled account of why they belong together (Ibid, cf. Pollock 1986: 183, Kvanvig 2003, Schroeder 2015). Our account of the phenomenon lacks the elegance and theoretical illumination we should want (cf. Zagzebski 2017). Some phenomena lack depth. Perhaps knowledge is one of these, though that would be surprising and a bit of a letdown.

Compare: good deeds are an interesting subject matter. Recognition is an interesting subject matter. It doesn't follow that good deeds that get recognized form a similarly interesting or philosophically important subject matter. If we had a branch of philosophy that concerned itself primarily with good deeds that get recognized, that would be surprising. We would expect that philosophers would either be able to state an interesting theoretical connection between the

[31] Cf. Putnam (1975)
[32] Kirkham (1984) embraces the consequence.
[33] One might object that the anti-Gettier condition is meant to be the connection between justification and truth, so it isn't a mere aggregate. My concern is that the connection just seems like another dimension of assessment entirely. I take it this is the worry that the comments about JTB+ analyses being *ad hoc* and unilluminating are gesturing toward.

goodness of a deed and its recognition, or the discipline would shift its focus. In the latter case, philosophers would say that the philosophically interesting thing is either the goodness of the deed or whether it gets recognized. We should expect the former. It seems like an essential property of the deed: it has to do with the excellence of the deed considered as a deed. The latter is an accidental relational property of the deed. It isn't clear why I should be interested in it at all insofar as I am interested in deeds.

Something similar happened in epistemology. There is a temptation to think of (1) as giving us the essential properties of reasons, in the sense that a reason is a good one just in case it rationalizes the subject's belief. There is a clear sense, after all, in which Gettier victims have good reasons for what they believe, despite their failure to have knowledge. (2) then appears to be an accidental relational property of the reason. This makes a phenomenon composed of (1) and (2) seem like an odd thing to have as the primary subject matter of an entire discipline. So, there is a tendency to say that epistemology is really about (1) and (2) is mentioned, if at all, as a codicil.

In *The Structure of Empirical Justification*, BonJour is giving a theory of the role reasons play in rationalizing belief. He briefly notes that we need more than truth and rationality for knowledge, but he brackets the point because the relation of this something more to justification/rationality is "obscure" to him (BonJour 1985: 5). So, he takes the theory of justification to be "the most fundamental sort of epistemological theorizing" (BonJour 2003: 40), the one that is interesting from a first-person perspective (BonJour 2003: 39). Fumerton also takes rationality to be the philosophically interesting (from a first-person perspective) "core of a concept of knowledge" (1995: 5ff, cf. Kaplan 1985 for the same argument), which makes the

philosophically less interesting periphery seem like an afterthought. Pollock and Cruz go further. Not only do they take (1) to be the philosophically interesting component of knowledge, they think that (2) makes knowledge a perverse concept:

> What the Gettier problem really shows is what a perverse concept knowledge is. One can do everything with complete epistemic propriety, and be right, and yet lack knowledge because of some accident about the way the world is. Why do we employ such a concept? (Pollock & Cruz 1999: 14)

This is given as a reason for focusing their epistemology book (curiously titled *Contemporary Theories of Knowledge*) on justification rather than knowledge. Epistemology is to be done from the first-person perspective and only the theory of justification is action-guiding (Pollock & Cruz 1999: 124). I submit that when so many of the leading figures in the theory of knowledge tell us that only one component of it is *the* philosophically interesting part, something has gone wrong. Yet this is what we should expect if the account they offer is gerrymandered. This leads me to think that we haven't really seen our way to clarity as to how the anti-Gettier condition is the *connection* between rationality/justification and truth. Despite knowing on some level that we can only understand what rationality is in terms of this connection, there is a tendency to abstract away from that relation when theorizing about it. We then end up (implicitly at least) thinking of it as a self-contained phenomenon and then face the difficult task of trying to glue it together with another self-contained phenomenon later.

2.3 Two Kinds of Good Reason

The basic idea I want to consider here is that we can make sense of the role reasons play in converting true belief into knowledge in terms of the role they play in rationalizing belief. We

can do this by requiring that when they convert true belief into knowledge, they rationalize the subject's belief in a way that does not depend on her ignorance of the facts. This seems like a step in the right direction.

This is the basic idea behind defeasibility solutions to the Gettier problem, of the sort considered in the last chapter. Robert Audi puts the point well here,

> [The conditions I propose are] are related to, and to some extent unified by, justification in the following way: if [the subject] *believed* that his circumstances were as they are in relation to the beliefs in question, he would not be justified in holding them. More generally, it is at least often the case that what undermines S's would-be knowledge without undermining his justification *would* undermine his justification if he believed it. And this makes it natural to say, though it does not require us to say, that much defeated knowledge is defeated because S's justification is in some way insecure or defective. (Audi 1993: 205)

In the chapter from which the quote is taken, Audi is talking about Gettier-style cases having to do with perceptual knowledge specifically. However, the same point can be made in the classic Gettier cases as well.

Audi puts the point in terms of justification, but it can easily be translated back into the terms I have used to frame the paper. Internalist justification is a matter of having sufficiently good reasons or adequate evidence. This comes to the same thing as the subject's belief being rationalized. Of course, a belief can be both rationalized and true yet fail to be knowledge. What goes wrong in such a case? Audi's point is that, at least often, the problem is that the reasons that

rationalize the subject's belief are defective or insecure. Their insecurity consists in the fact that, were the subject to believe that her circumstances are as they really are, that reason would no longer be able to play its rationalizing role for her.

Essentially the same point is made by a number of other authors. Sometimes instead of talking about a subject's justifying reasons being defective or insecure they say that in the kind of case Audi is imagining those reasons are "rotten" (de Almeida & Fett 2016), "not worth having" (Klein 2017), or they don't "epistemize" (Steup 1996: Chapter 1). Similarly, when the subject knows her justifying reasons are "robust" (Schroeder 2015), "truth-resistant" (Moser 1988, 1989) or "defeat-proof" (Klein 2017). The important point for our purposes is how the reason's merits or defects are understood in terms of whether it could rationalize the belief if she were to know the facts about her situation.

Suppose a student tells the professor that he owns a Ferrari, and the professor forms the belief that someone in her class owns a Ferrari (cf. Lehrer 1965). It turns out that someone in her class does own a Ferrari, but it is not the student who told her he did. That student is a pathological liar, although the professor had no reason to suspect this. The professor has a true belief that is rationalized by the testimony she has received. It nonetheless falls short of knowledge because it wouldn't rationalize her belief if she were aware that he is a pathological liar.

Now we have an explanatory connection between the role that reasons play in converting true belief into knowledge and the role they play in rationalizing beliefs. This makes some progress toward addressing the gerrymandering worry, since the absence of any kind of

interesting explanatory connections between the components of a theory are what makes the theory gerrymandered. Although this is a good start, I worry that more needs to be said.

First, the point is made as a counterfactual. If learning the facts about her situation would prevent the reason that rationalizes her belief from continuing to do so, then her reason does not turn her true belief into knowledge. Why does this counterfactual bear on her *actual* epistemic status? We can all agree that her reason *wouldn't* rationalize her belief if she were to know more. But she doesn't know more. Why is it the case that her reasons *are* defective? Why isn't it just that they would be if she were to know more? After all, as things stand, they rationalize her belief. Indeed, it is a little strange to say that her justification is defective, rotten, etc. Considered as justification, it seems fine. She is just as justified as the knowing subject, otherwise Gettier cases wouldn't be counterexamples to JTB analyses. The same point can be translated into reasons-talk. If we understand what it is to be a good reason the way we have been, then her reasons aren't defective, rotten, etc. They do exactly what they are supposed to do: rationalize her belief relative to the total evidence she in fact has. So, although there is an explanatory connection between (1) and (2), it isn't of the right sort.[34]

There are other problems with putting the point in counterfactual terms. There are many ways to make the antecedent of a counterfactual true and some of them have unintended consequences (for an extended discussion of this point in relation to the Gettier problem, see Shope 1983). For this reason, we should make essentially the same point slightly differently.

[34] It is sometimes argued that knowledge is more explanatorily powerful than mere rational true belief because it is more robust in the fact of new evidence and hence less likely to be lost after one learns more (Williamson 2000: Chapters 2,3). However, rational beliefs with no factual defeaters are even more robust to new evidence, but knowledge is still more interesting.

Instead of stating the point in terms of whether the subject's reasons would rationalize her belief were she to know the facts, they state it in terms of whether there are facts that either trump or mitigate her reasons. The key idea here is that in order for the subject to have knowledge, her justifying reasons must be good in both the subjective normative and the objective normative sense. That is, her reasons must be good assessed relative to her total evidence and also good assessed relative to the facts. In Lehrer's Ferrari example, the subject's reasons are good insofar as we assess them relative to her total evidence (i.e., she has a subjective normative reason). However, they are not good insofar as we assess them relative to the facts (i.e., she does not have an objective normative reason). Her testimonial reason is mitigated (or undercut) by the fact that the testimony came from a pathological liar.

What I say here won't depend on any particular account of the necessary and sufficient conditions for reasons being good in either sense. There is a cluster of very similar proposals in the defeasible reasoning tradition.[35] More recently, John , (2007; 2012, cf. Schroeder 2021: 92 ff.) has developed a formal model of defeasible reasoning using default logic. Similarly, Mark Schroeder (2007; cf. 2015) has given a recursive specification of how reasons are weighted that can be applied to either subjective normative or objective normative reasons. The differences between these views needn't concern us here. I am interested in the genus of which they are species.

Each species gives us an account of which conditions must obtain for a reason to be good in the subjective normative and the objective normative sense. Even after we have determined

[35] See fn. 12.

which conditions must obtain for the subject's reason to be good in both senses, we still need to determine what it is for her reason to be good in both senses. Each species of the genus gives us an account of which conditions must obtain for the subject to have a subjective/objective normative reason. They do this by giving an account of how the subject's justifying reasons need to stack up against a body of potentially countervailing reasons to be sufficient. In the case of subjective normative reasons, the body of potentially countervailing reasons is the subject's total evidence. In the case of objective normative reasons, the body of potentially countervailing reasons is the totality of the facts.

However, this doesn't tell us what it is to have a good reason in either sense. To see the point, consider the following example. Suppose you can determine whether you have high cholesterol by spitting on a stick and putting it in chemical solution A and you can determine if you have lupus by spitting on that same stick and placing it in chemical solution B. Does this show that there is an interesting connection between what it is to have lupus and what it is to have high cholesterol? No.

Similarly, it may be the case that you can determine whether a subject has a subjective normative reason by performing a procedure relating her reason to one body of evidence and you can determine if she has an objective normative reason by performing the same procedure relative to the totality of the facts. But this doesn't show that subjective normative reasons and objective normative reasons have any more explanatory connection than lupus and high cholesterol would have if chemical solutions A and B were actual. Because he recognizes this point, Schroeder offers the following: a subjective normative reason is a reason that would be objective normative if the subject's beliefs were true (2007: 14; 2008; 2015).

Some counterfactual in the neighborhood is true. I prefer the following: if the world were to be as the subject justifiably takes it to be, then her subjective normative reasons would be objective normative reasons.[36] Either way, the counterfactual doesn't explain what needs to be explained. As an account of the conditions that must obtain for reasons to be subjective normative, this is fine. However, it doesn't tell us what it is to be a subjective normative reason. It doesn't give an adequate account of their conceptual connection with objective normative reasons. As with Audi's proposal, the problem stems from the fact that the analysis is stated as a counterfactual (though this time not for the reasons identified by Shope 1983). If the subject's beliefs were true, the subjective normative reason would be an objective normative reason (mutatis mutandis for my proposed revision). This tells us nothing about the standing of her reasons in the actual world. It tells us nothing about how subjective normative reasons are connected to objective normative reasons in the actual world. If wishes were horses, then beggars would ride. Wishes aren't horses. So, what is the connection between wishes and horses in the actual world? There isn't one! If the subject's beliefs were true, then her subjective normative reasons would be objective normative reasons. What about when they aren't true? What is the connection between subjective and objective normative reasons then? Schroeder's account doesn't help us here and neither does any other counterfactual account. Perhaps this is why in more recent work he remains agnostic about the connection between the two (Schroeder 2021: 45).[37] But there surely is still a connection between subjective normative reasons and objective

[36] Putting it this way runs afoul of the reasons-first order of explanation. There is a more cumbersome way of making the same point that does not: if the world were as the subject rationally takes it to be (where rationality is determined by the balance of subjective normative reasons), then her subjective normative reason would be an objective normative reason.

[37] Though see (Schroeder 2021: 69) for a possible lapse into his earlier view.

normative reasons in the actual world, even when the subject's beliefs are false, and the connection isn't identity. Surely there is a connection *from the subject's point of view*.

This is what explains why Schroeder's counterfactual is true. If the subject's beliefs were true, her subjective normative reason would be an objective normative reason. This is because the subject's beliefs are the facts, so far as she can tell. Similarly, her justified beliefs (i.e., those for which she has sufficient subjective normative reasons) are the facts, so far as someone with her total evidence can tell.[38] So, her subjective normative reasons for thinking that those are the facts are just objective normative reasons, so far as one with her total evidence can tell. If the facts are as she takes them to be, then her subjective normative reasons are not misleading. In that case, they are objective normative reasons. But even if they aren't in fact objective normative reasons because her total evidence is misleading, it remains the case that her subjective normative reasons appear to be objective normative reasons to one with her total evidence. That is what it is to be a subjective normative reason and it is what explains why the counterfactual is true. In general, the counterfactual is explained in terms of the actual. This is no exception.

Of course, more needs to be said about the sense of "appear" at play here, and more will be said in the next section. The key point is that we don't have a non-gerrymandered account of knowledge until we can articulate the connection between what is to be a subjective normative reason and what it is to be an objective normative reason. None of the species mentioned above do that. They give us an account of which conditions must obtain for reasons to be normative in

[38] Strictly speaking, these are her justifiable beliefs. I ignore the subtlety here to streamline exposition.

both senses. But they don't give an account of what it is to be normative in either sense that makes clear the connection between the two senses of "normative reason" because they don't tell us what it is to be a subjective normative reason in a way that explains the connection.

2.4 Good Reasons are Apparent to the Knowing Subject

I will begin by offering some further motivation for the claim that what it is to be a subjective normative reason is to be an apparent objective normative reason to one in the subject's epistemic position (i.e., with her total evidence). It seems clear that there is an important sense in which the subject in a Gettier case has good reasons for her Gettiered belief (the same goes for justified false beliefs). Reasons, by their very nature, rationalize belief. This makes their role in rationalizing belief essential to them. When reasons rationalize belief, they do something they are supposed to do *qua* reason. That makes them good reasons, at least in one important sense. However, this is not all that reasons are supposed to do. In addition to standing up to scrutiny in light of the subject's possessed evidence, they are also supposed to stand up to scrutiny in light of the facts. When reasons don't stand up to scrutiny in light of the facts, they fail to do something they are supposed to do *qua* reasons. For this reason, it would make perfect sense for an enlightened third-party to say the following about a Gettier subject, "Her reasons are no good, although she is in no position to tell." The third-party might qualify the remark by noting that the Gettier subject's reason was good *for her* at the time. That is, it was not good in the objective normative sense but perfectly fine in the subjective normative sense.

These remarks indicate not only that there are two senses of "good reason" or "normative reason", but also gesture in the direction of how they relate. In the last section, I said that what it is to be a subjective normative reason is to be apparently good in the objective normative sense.

As we just saw, it would be quite natural to say that a Gettier subject's reasons weren't any good, although she was in no position to tell. That is, she didn't have an objective normative reason, although she reasonably thought otherwise. I say this is why she had a subjective normative reason. To have a subjective normative reason just is to apparently have an objective normative reason. This explains why the remark from the enlightened third party makes sense.

I don't mean that for a reason to be good in the subjective normative sense the subject must also have a higher-order reason to think that it is good in the objective normative sense.[39] If this seems puzzling, the following remark may help. The question of what to believe is transparent to the question of what is the case (cf. Evans 1982, Moran 2001). This does not mean that it is impossible to believe what is not the case. It just means that when you are trying to figure out what to believe, you do so by trying to figure out what is the case. I urge that a similar point applies to figuring out what your reasons are. When you are trying to figure out if your perceptual evidence is really a reason for you to form a belief about the address of the house you just drove past, you try to determine whether there is in fact some reason to doubt your senses in this situation, i.e., an objective normative reason. You don't first figure out the subjective normative reasons and then try to build a bridge to the objectively good reasons through meta-reasons. Rather, you just figure out what the objective normative reasons are. The fact that this takes place within the inquirer's limited perspective does not mean that they first take stock of how things are from within that perspective and then try to see how it lines up with the facts (*pace* McClelland & Chihara 1985: 81; Fumerton 1995: 174).

[39] Thanks to an anonymous referee for making clear the importance of driving this point home.

In fact, Fumerton, BonJour, and Pollock & Cruz's remarks from earlier have things backwards. They think that from the first-person perspective you check to see if your reasons are good in the subjective normative sense. Then there is no further inquiry into how they line up with the facts. You can't jettison your perspective to go check. So, they think the theory of good reasons in the subjective normative sense (i.e., justification as they understand it) is the heart of epistemology. How reasons line up with the facts is just something you have to tip your hat to when knowledge comes up.

I urge that what you care about from the first-person perspective is how your reasons line up with the facts, i.e., goodness in the objective normative sense (cf. Littlejohn 2012a). Once you've determined that as well as someone in your position can, there is no further question about your subjective normative reasons. The goodness of reasons in the objective normative sense is the heart of epistemology. Goodness in the subjective normative sense is just a finite subject's best estimate of that. There is nothing left over to which we must reluctantly tip our hats.

A good reason in the subjective normative sense is apparently (to one in the subject's position) a good reason in the objective normative sense. To see this, try to consider a subject for whom R is a subjective normative reason, i.e., it rationalizes some belief for her, but for whom it is also reasonable to believe that, if she were better informed, R would no longer be able to rationalize that belief for her. You can't conceive of such a subject. If the second conjunct is true, the first isn't. If it is reasonable for her to think that R only rationalizes her beliefs because she is

in a benighted position, then R isn't really a good reason for belief to one in her position.[40] So, it doesn't really rationalize her beliefs. It is a merely *prima facie* reason defeated by higher-order evidence.

Conversely, the more reasonable it is to think that your reasons will continue to support your belief relative to more comprehensive bodies of evidence, the weightier they become *qua* subjective normative reasons. The professor in Lehrer's Gettier case was reasonable in taking the student's testimony to be a good reason in the objective normative sense. Rarely does testimony about car ownership fold under further inquiry. However, in this case it did. So, it was not a good reason in the objective normative sense, but it was a good reason in the subjective normative sense. This is because it was reasonable for the subject to take it as good in the objective normative sense. If she had even more reason to think that the testimonial reason would survive further inquiry (e.g., misleading evidence that the student is more honest than average), then her reasons would be even better subjective normative reasons.

The key takeaway is that my account does not make subjective normative reasons harder to come by due to their identification with apparent objective normative reasons. I am not saying that having a subjective normative reason requires the absence of countervailing reasons in the subject's possessed evidence and, *additionally,* reasons to think there aren't countervailing reasons out there of which she is presently unaware. Rather, by identifying subjective normative reasons with apparent objective normative reasons, I am saying that the absence of countervailing reasons in the subject's possessed evidence *just is* apparent lack of sufficiently

[40] Ballantyne (2015) makes much of this point.

weighty countervailing reasons “out there”. This is the upshot of the transparency point. The subject’s beliefs are her take on the facts and her justifying reasons are her take on the reasons that will ultimately stand up to scrutiny in light of the facts.

This explains why “Custer died at Little Big Horn, but I don’t know that he did” sounds bad (Littlejohn 2012a: 2; cf. Smithies 2012). You can’t rationally believe the conjunction. If you rationally believe the first conjunct, then you are committed to the claim that your reasons are of the sort that turn rational true beliefs into knowledge, i.e., objective normative reasons. You are committed to this because you take yourself to have a subjective normative reason (otherwise asserting the first conjunct would be bad), and that is just an apparent objective normative reason. But if your apparent objective normative reason really is as it appears to be, then you know the first conjunct.

This does not generate a regress. When the subject uses r as a reason for believing p by basing belief that p on r, she implicitly regards r as a good (objective normative) reason for believing p. This is the transparency point, with the extra note that one’s attitude toward one’s reasons can be implicit through basing. She does so reasonably if, so far as one can tell from her perspective, r is a good reason for p. When this condition obtains, it will sometimes be because she has some further reason for so regarding r. In other cases, r will be a *prima facie* reason for believing p. That is, it will not depend positively on other reasons for its own status as a good reason. Rather, it depends negatively on them: It doesn’t depend on them for support, only the absence of defeat. This halts the regress. I won’t argue further for the legitimacy of the distinction between positive and negative epistemic dependence until Chapter 4.

The distinction between positive and negative epistemic dependence also applies to your attitudes towards your reasons. Your belief that you have hands when you see them in ostensibly normal conditions depends negatively on the absence of reasons to think your perceptual states are misleading in this instance. It does not depend positively on independent reasons to think they are not illusory. Similarly, it is reasonable to take your perceptual reason to be an objective normative reason in the absence of reasons to think otherwise. If it is not reasonable to take your reason to be an objective normative reason, then you cannot reasonably believe on its basis. If you are not in a position to reasonably believe that your reasons will withstand scrutiny in light of the facts, then they aren't even good reasons relative to your total evidence (i.e., they are not even subjective normative reasons). Why is it reasonable for the Professor in Lehrer's example to believe that the student owns a Ferrari? Because he said so, and it is reasonable for her to believe that her testimonial reason is not trumped or mitigated by further facts. Testimony of this sort usually holds up. If it weren't reasonable to think the reason will hold up in the fact of the facts, then she would have no business believing on its basis in the first place. She should suspend judgment instead.

The problem with which we began is that the connection between (1) and (2) is unclear. Without a clear account of the connection, the concept KNOWLEDGE appears gerrymandered. We have now seen how to avoid this result. Reasons that do (1) are (by their very nature) reasons that apparently do (2). We can now say a bit more to alleviate the gerrymandering worry. In the last section we saw a family of approaches according to which (1) is a matter of how the subject's reasons square with her possessed evidence and (2) is a matter of how they square with the facts. This helps to the extent that it shows (1) and (2) to be complementary perspectives on a

single thing: the subject's reasons. However, the gerrymandering worry still rears its head. Unless more is said, it appears that when we assess the subject's reasons relative to her possessed evidence, we are checking for one property: rationality. When we assess her reasons relative to the facts, it isn't immediately clear that we are looking at that property from a different perspective as opposed to looking at an entirely different property. Schroeder, for instance, calls the dimension of assessment relevant to (2) "alethic" (2015). This gives the impression that we are using one procedure to check for two very different things: one alethic and the other properly epistemic. This is the point made earlier by the lupus/cholesterol example. We saw that Schroeder's counterfactual analysis of subjective reasons in terms of objective reasons doesn't help matters.

Notice, however, that in assessing the subject's reasons relative to the totality of the facts, we are treating the totality of the facts like a body of evidence, like that possessed by any finite subject at any given time. The body of evidence composed of all and only the facts is the (regulatory) ideal end point of inquiry. It is the point at which there is nothing further into which one can inquire. If the subject's reasons square with the totality of the facts (according to the procedure laid down by whichever species of the genus considered in the last section proves to be correct) then her reasons are good relative to the perspective that is the ideal end point of inquiry. This enables us to better see the explanatory connection between (1) and (2). Assessing the subject's reasons relative to her own perspective (i.e., total evidence) and the perspective composed of all and only the facts is not like dipping her beliefs in unrelated chemical solutions. Rather, it is to assess the subject's reasons relative to her own perspective and also relative to the ideal end point of the activity in which she is engaged as she reasons.

The view offered here can, with a slight emendation, also address another problem for Schroeder's view (indeed, other species of the genus as well). It is possible for reasons to be both subjectively and objectively sufficient, and a belief to be based on them, but that belief might still fail to be knowledge. I diagnose the problem in the following way: it is possible for good reasons to be apparent to a subject without the goodness of those reasons being apparent to her. The latter is necessary for knowledge.

This point is illuminated by John Turri's (2010) example in which the subject justifiably believes that p, and that if p, then q. Those are the reasons for which she believes q. These are good reasons (in both the subjective and objective normative senses) to believe that q. However, she doesn't know that q. Instead of inferring q from those reasons using *modus ponens*, she infers it using the inference rule: infer q from any two premises. The reasons themselves are perfectly fine, the subject's use of those reasons is problematic (Ibid).

On my account, good reasons are apparent to this subject, but the goodness of those reasons is not. What makes those reasons good is the fact that they are fit to figure in a modus ponens argument. One needn't form an explicit belief to suitably appreciate this fact. One suitably appreciates this fact just by basing one's belief that q on p and p → q by inferring q from these premises in accordance with modus ponens. By inferring one's belief that q in this way, one suitably appreciates the goodness of one's reasons for believing q by appreciating what it is about one's reasons that makes them good reasons to believe q: their role in modus ponens. That is, it is not just apparent that these reasons are good: what makes them good is apparent as well. So, strictly speaking, the goodness of good reasons is apparent to the knowing subject. That is a mouthful, so I will stick with the punchier "good reasons are apparent to the knowing subject" as

the mantra of the paper. However, the official position is nonetheless that their goodness must also be apparent.

We can now state what our interest in knowledge is all about. We are interested in whether good reasons are apparent to the subject. We are interested in whether her reasons really are as they appear to her. Things really being as they appear to one is clearly valuable (cf. Williamson 2000: 40). (2) is a matter of whether the subject's reasons really are good and (1) is a matter of whether they are apparently good. They are both about the goodness of reasons. Interest in their apparent goodness just is an interest in their goodness, from the point of view of a finite and possibly mistaken perspective.

I should also mention that the subject having good reasons in the objective normative sense is truth-entailing, so there is no need to mention it separately.[41] The account I recommend is still fallibilist. The fallibilist maintains that we can know on the basis of defeasible reasons. I agree. The knower's reasons are defeasible, they are just not, in fact, defeated (in the same way my books are flammable but not, in fact, on fire).

When you believe something, you know it just in case your belief is based on good reasons apparent to you. This sounds so obvious that a non-philosopher would think it is hardly worth stating. This is a point in favor of the view. It shows us how to say what we were inclined to say before we entered the epistemology classroom in a way that is consistent with the lessons epistemology has taught us.

[41] Rosenberg (2002) has a proposal that is similar in this respect, although relativist in a way mine isn't.

Furthermore, the metaepistemological view that emerges doesn't require us to say that epistemology is really about one aspect of knowledge rather than the other. Epistemology is about reasons to believe (i.e., our metaepistemology is "reasons-first"). Epistemic assessment is about the goodness of those reasons. Once we see what knowledge is, we see that it is the unified, interesting phenomenon we implicitly take it to be. We have made sense of knowledge in terms of a kind of assessment that is an important part of how we understand ourselves. The metaepistemological view I recommend enables us to see epistemology as being unified by a single kind of assessment, one everyone implicitly understands before entering the epistemology classroom.

2.5 Conclusion

I have argued that the role reasons play in rationalizing belief and the role they play in turning true belief into knowledge can only be understood in tandem. The picture that emerged is that a subject knows just in case her belief is based on good reasons apparent to her. This enables us to make sense of and reflectively endorse our practice of trafficking in knowledge attributions. What appears to us as a unified and significant aspect of our lives is what it appears to be. Certain tendencies in epistemology make it difficult to see this. Those tendencies are difficult to avoid. Once we see that the components of knowledge can come apart, it is difficult to see what unites them. Recognition of this difficulty, even if just inchoate, can incline us to think that the Gettier intuition is faulty (Weatherson 2003, Olsson 2015, Weinberg, 2017), or that knowledge shouldn't be analyzed in terms of rationality plus something else (e.g., Zagzebski 1994, 2017), or perhaps that we shouldn't try to analyze it at all (Williamson 2000). I hope to have shown that we can understand knowledge in terms of a prior understanding of reasons without our analysis turning into the "*ad hoc* sprawl" (Williamson 2000: 31) that gives these alternatives much of

their appeal. There is nothing *ad hoc* about the phenomenon of one's reasons really being as they appear.

We can now state what our interest in knowledge is all about. We are interested in whether good reasons are apparent to the subject. We are interested in whether her reasons really are as they appear to her. Things really being as they appear to one is clearly valuable (cf. Williamson 2000: 40). (2) is a matter of whether the subject's reasons really are good and (1) is a matter of whether they are apparently good. They are both about the goodness of reasons. Interest in their apparent goodness just is an interest in their goodness, from the point of view of a finite and possibly mistaken perspective.

I should also mention that the subject having good reasons in the objective normative sense is truth-entailing, so there is no need to mention it separately.[i] The account I recommend is still fallibilist. The fallibilist maintains that we can know on the basis of defeasible reasons. I agree. The knower's reasons are defeasible, they are just not, in fact, defeated (in the same way my books are flammable but not, in fact, on fire).

When you believe something, you know it just in case your belief is based on good reasons apparent to you. This sounds so obvious that a non-philosopher would think it is hardly worth stating. This is a point in favor of the view. It shows us how to say what we were inclined to say before we entered the epistemology classroom in a way that is consistent with the lessons epistemology has taught us.

Furthermore, the metaepistemological view that emerges doesn't require us to say that epistemology is really about one aspect of knowledge rather than the other. Epistemology is

about reasons to believe (i.e., our metaepistemology is “reasons-first”). Epistemic assessment is about the goodness of those reasons. Once we see what knowledge is, we see that it is the unified, interesting phenomenon we implicitly take it to be. We have made sense of knowledge in terms of a kind of assessment that is an important part of how we understand ourselves. The metaepistemological view I recommend enables us to see epistemology as being unified by a single kind of assessment, one everyone implicitly understands before entering the epistemology classroom.

Chapter 3: First-Class & Coach-Class Knowledge

Here is a familiar idea from the defeasible reasoning literature: your reasons[42] need to pass muster along two dimensions of assessment for you to have knowledge. The first dimension requires that your *prima facie* reasons not be defeated by your total evidence. Otherwise, your belief is irrational or unjustified. But they also can't be defeated by facts of which you are unaware. Factual defeat can't make the belief irrational or threaten justification. However, it can threaten knowledge. If your justification is factually defeated, then it is (at best) accidentally true, no matter how rationally held. This is the upshot of Gettier cases. Call the first, rationality-involving dimension of assessment "subjective normative" and the other fact-involving one "objective normative". There is a fact of the matter about whether a subject's reasons are good along either dimension, the fact is just determined by different things in each case. It is determined by the subject's total evidence in the former and the totality of the facts in the second.

If this is right, then we can understand why Gettiered subjects have accidentally true beliefs. Non-accidentally true belief requires that the subject's reasons pass muster along both dimensions, but Gettier subjects have reasons that only do well along the former. This proposal is familiar.[43] I will argue here that it can also be used to explain graded accidentality. That is, some beliefs are neither straightforwardly knowledge nor non-knowledge. This is not (in the cases I'm thinking of) because "knowledge" is vague. Rather, it is because the subject's reasons

[42] I will generally talk about reasons since I am a reasons-firster. However, the point I want to make here can be put in terms of evidence if you prefer (cf. Klein 2017). Some locutions (e.g., "total evidence") sound better when I speak of evidence instead of reasons. The difference is only stylistic.

[43] See Lehrer (1965, 1970), Lehrer & Paxson (1968), Hilpinen (1971), Johnsen (1974), Swain (1974), Barker (1976), Klein (1971, 1976, 1980), Pollock (1986), Moser (1989), Audi (1993), Schroeder (2015), de Almeida & Fett (2016). See Shope (1983) for a useful overview.

pass muster along the first dimension of assessment full-stop but the latter dimension only with certain qualifications. Distinguishing different qualifications, then, allows the discovery of different ways that non-accidentally true belief can be graded.[44] I will argue that in each case we have instances of "coach-class" rather than "first-class" knowledge.

Accidental truth comes in degrees. First-class knowledge is completely non-accidentally true. It is the limiting point on the accidentality spectrum. Coach-class knowledge includes all cases where there is some accidentality involved, but not enough to make it a Gettier case. So, coach-class knowledge admits of differences in degree. There are also differences of kind among cases of coach-class knowledge. You get coach-class knowledge when your reasons are ultimately good relative to certain qualifications but not full-stop. Different qualifications give rise to different species of coach-class knowledge. Indeed, we will see a different species of the coach-class genus in each section. All species of the genus share a structural property: the subject's justifying reasons are ultimately good insofar as x but not full stop. Different values of x give us different species. Although I will stick with the "first-class/coach-class" terminology, we could just as well call it "unqualified vs. qualified" knowledge.

Apart from providing an interesting new kind of case, this paper will also give us an ecumenical way of settling the disputes about knowledge from falsehood and fake barns. In these cases, we can split the difference in a principled way between the parties to the dispute. In the final section of the paper, I will briefly argue that a safety-based analysis of graded non-accidental truth doesn't fare as well as the account I recommend. I leave open the possibility that

[44] The way I grade knowledge is distinct from, but compatible with, Reed's (2013) way of doing it.

a virtue-theoretic account can handle the cases of interest, although I won't go into the matter here.[45]

In section (I) I will briefly go over the aspects of the defeasible reasoning tradition that matter for my proposal. In section (II) I will consider a new kind of case in which the subject's reasons are not ultimately good considered in themselves, but they are a proper part of the ultimately good reasons in the objective normative sense. Insofar as we consider them as a proper part of those reasons, they are ultimately good. In section (III) I do the same thing except by focusing on cases where the subject's reasons are epistemically similar to the ultimately good reasons. I argue that cases of knowledge from falsehood are like this. In section (IV) I argue that intuitions about fake barns are less firm than those about classic Gettier cases because we must qualify our answer to the question whether their justifying reasons are ultimately good by distinguishing their foreground reasons from their background assumptions. The former are ultimately good while the latter are not. In section (V) I respond to objections and give reasons to choose the account developed here over a counterfactual analysis of graded non-accidentally true belief.

3.1 Background

In this section I will describe the core commitments of the defeasible reasoning tradition[46] before going on to apply it in the following sections.

[45] See Carter (2014) for what I take to be the most promising virtue-theoretic approach to a similar phenomenon.
[46] Some prominent objections to this approach are those of Feldman (2003), Foley (2012) and Turri (2012). In my view, an ample response to all three can be found in de Almeida & Fett (2015). Lasonen-Aarnio (2010 a,b, 2014) and Hytch & Benton (2016) raise a distinct set of worries I won't be able to address in this paper.

A necessary condition for a subject to know p is that they believe p on the basis of a *prima facie* good reason. *Prima facie* reasons are foundational reasons. This doesn't mean that beliefs formed on their basis are infallible or incorrigible. A moderate foundationalist will hold that *prima facie* reasons don't depend positively on other reasons for support, but nonetheless depend negatively on the absence of countervailing reasons.[47] These countervailing reasons are called "defeaters". Defeaters are the mirror image of *prima facie* reasons. *Prima facie* reasons count provisionally in favor of believing something and defeaters count provisionally against believing it. For example, perceptual experience as of a red wall is a *prima facie* reason to believe that the wall is red. It is an all-things-considered reason to so believe just in case it is undefeated. If the subject were to be told that the lighting in the room is non-standard, for example, they would have a defeater. Nothing I say in this paper will depend on any particular analysis of defeaters. This is a virtue, not a lacuna. The intuitive idea is that they are *prima facie* reasons to give up a belief (cf. Ballantyne 2015). There are many ways of trying to make the intuitive idea more precise, but the details don't matter for my purposes here.[48] The things I say here are independently plausible and a complete account of defeaters should strive to accommodate them. In the last chapter, I will say a bit more about the metaphysics of reasons generally. However, the claims I make here don't depend on the claims I make there.

Since defeaters are *prima facie* rather than all-things-considered reasons to give up a belief, your *prima facie* reason can have a defeater but still be a good reason all-things-considered. This can happen when the defeater is itself defeated. For example, if you were to

[47] The first person to put it this way was Robert Audi (1993), although the idea is latent in earlier work on defeasible reasoning.
[48] See for Pollock (1986) for one influential account. For recent work, see Brown & Simion (2021).

learn that the person who told you the lighting is non-standard is a pathological liar, then the defeater would be defeated, and your *prima facie* reason would be restored. Not all defeater-defeaters do this, however. Some defeater-defeaters give you new reasons rather than restoring your (actual) justifying reasons.[49] For instance, suppose you were to learn that the lighting on the wall really is non-standard but you were also told that someone painted the wall red so it would still look the same even if the lighting were to change. The information about the painting of the wall defeats the defeater having to do with the lighting. However, it doesn't restore your justifying reason (i.e., your perceptual experience). Despite the new information, your eyes are still not to be trusted. Your perceptual experience is now replaced, rather than restored, by the testimony about the painting of the wall. Going forward, I will call defeater-defeaters that do not restore the subject's justifying reasons "non-restoring defeaters"[50] and those that do restore their justifying reasons "restoring defeaters".

When we want to know whether a subject's belief is justified, we check to see if she has *prima facie* reason for that belief. If so, we check for defeaters in her possessed evidence. Non-restoring defeaters in her possessed evidence defeat her justification, so I will call them "justificational defeaters". But a belief can be justified yet fail to be knowledge. This is what we see in Gettier cases. In these cases, the subject's justification is defeated by a fact of which she is unaware. So, I will call these "factual defeaters".[51] The key insight of the defeasible reasoning tradition is that knowledge, over and above true belief, is a matter of the standing of the subject's reasons and to determine their standing we must look at those reasons from the subject's

[49] Cf. Klein (1980), Pollock (1986), de Almeida & Fett (2016).
[50] I get the term from de Almeida & Fett (2015). Similar terminology can be found in Klein (1980).
[51] Some call them "knowledge defeaters" (Audi 1993) or "propositional defeaters" (Bergmann 2006).

perspective but also *sub specie aeternitatis*. The subject's reasons must be good in the subjective normative sense which requires that her justifying reasons not have a non-restoring defeater in her possessed evidence. It also requires that they be good in the objective normative sense which consists in that justifying reasons not have a non-restoring factual defeater.

Peter Unger (1968) defined knowledge as non-accidentally true belief. Everyone seems to agree to this much, although there is much disagreement about what it is for a belief to be non-accidentally true. Some deny that we can understand the kind of accidents that undermine knowledge without a prior understanding of knowledge itself. These people are "knowledge-firsters".[52] Others think we can give a modal analysis of the kind of accidents that undermine knowledge that will help explain knowledge.[53] This is what safety theorists propose. Those working in the defeasible reasoning tradition, alternatively, think we can understand the kind of accident that undermines knowledge in terms of *prima facie* reasons and defeaters.[54]

Importantly, both dimensions of epistemic assessment are explanatorily prior to knowledge and neither requires a modal analysis. The key components of the account are *prima facie* reasons and defeaters. *Prima facie* reasons are epistemic foundations. Proponents of the defeasible reasoning tradition are typically internalists about justification, which is the view that I recommend. I won't try to give a definition of internalism here, as there may be more than one internalism/externalism controversy. At the very least, however, internalists are committed to explaining the epistemic potency of the foundations in non-modal terms. There are a few options

[52] See Williamson (2000), Sutton (2007), Bird (2007), Reynolds (2013). Some combine knowledge-first and virtue epistemologies, e.g., Miracchi (2015), Kelp (2016), Millar (2019).
[53] The version of this approach I will give the most attention is Pritchard (2005). See also Hawthorne (2004), Pritchard (2012) and Hawthorne & Rabinowitz (2017) for different ways of pursuing it.
[54] See footnote 2 for representatives of this tradition. Klein (2017) calls these "quality of evidence" views.

here. The epistemic potency might not be further explicable[55], it might be explained in terms of a basic entitlement to believe that things are as they seem[56], it might be explained in terms of the acquaintance relation[57], or perhaps some other way. I will remain neutral here, although I will have more to say about this in the next chapter. Defeat isn't given a modal analysis either. Defeaters are either part of the subject's possessed evidence or actual world facts.

Now that we know how the defeasible reasoning tradition explains non-accidentally true belief, I will go on to discuss how it can be extended to explain graded non-accidentally true belief. In the final section, I will argue that it does better than a modal account could. A quick terminological note: when I talk of reasons being ultimately good/bad ultimately defeated/undefeated in this paper, I mean in the objective normative sense of those terms, unless otherwise noted.

3.2 Noir-Style Cases

The first kind of case I will consider I am going to call "noir-style cases" for lack of a better term. The basic idea is familiar enough from old movies and contemporary network police procedurals. There is a detective investigating something. They come up with a hypothesis about what happened. They later come across countervailing evidence suggesting that their initial theory was incorrect. After that, they come across even more evidence suggesting that their initial theory was not so much mistaken as incomplete. The true account of what happened ends up being a more convoluted version of what they initially thought. Raymond Chandler's *The Big*

[55] See Pryor (2001).
[56] See Huemer (2007).
[57] See Fumerton (1995).

Sleep patterns this way, although the plot is too complicated to discuss it here. Here is an outline of a simpler case,

> **The Big Noir:** Marlowe was investigating the murder of James Cagney. Cagney was a known criminal and associate of crime boss John Huston. After asking around, Marlowe learned that Cagney was seen on more than one occasion leaving a hotel room with Huston's wife Ingrid Bergman. At the end of the first day on the case, Marlowe formed the belief that Huston killed Cagney. He figured that Bergman and Cagney were having an affair and Huston killed Cagney when he learned about it. His reason was that this gives a cohesive explanation of the available evidence. The next day while looking around the hotel room Marlowe learned that it was used for a counterfeit operation. After asking around some more, he learned that Huston had been defrauded of money he had tied up in a counterfeit operation. Over the next couple of days other associates of Huston started appearing in dumpsters all over the city. Marlowe began to doubt his initial hypothesis. Kirk Douglas, Huston's highest-ranking associate, then turned himself in and confessed to Marlowe that he, Cagney, Bergman and others had tried to defraud Huston of the money he had tied up in the counterfeit operation. During this time Bergman and Cagney began having an affair. Huston followed them to the hotel when he began to suspect Bergman of infidelity. When his suspicion of infidelity was confirmed, he also learned about the counterfeit scam. He killed Cagney because of all he had learned and then went after all his associates that he suspected were in on the scam.

Marlowe's initial reason for thinking Huston did it was that the jealous husband narrative gave a cohesive explanation of the available evidence. Once more evidence came in, the explanation

became less cohesive. Was the counterfeit somehow unrelated to the affair despite taking place in the same room? Were the killings of the other associates unrelated to the Cagney killing despite the fact that both were involved in the counterfeit scam? Once the last of the evidence came in the form of the confession, a new explanation became available. Marlowe's original explanation was a proper part of it. Insofar as Marlowe's original explanation is a part of the final explanation of the evidence, it offers ultimately good reasons for thinking Huston did it. Considered as a complete account of what happened, it does not offer ultimately good reasons for thinking that. Considered as a complete account it is defeated by the evidence it does not explain.

When Marlowe formed the belief that Huston did it at the end of the first day, there were plenty of factual defeaters. They did not defeat his justification because they were not part of his possessed evidence. They are like the defeaters we find in Gettier cases in this respect. However, it is less clear in this case than a classic Gettier case that Marlowe lacks knowledge. I submit that at the end of the first day on the case, his belief that Huston did it was somewhat non-accidentally true. This is because his reasons are ultimately good, subject to certain qualifications. We feel uneasy saying that they are ultimately good, but we also feel uneasy saying that they are not. They aren't as straightforwardly ultimately defeated as the Gettier victim's reasons. They are also not as straightforwardly ultimately good as the reasons a subject has when we have no qualms about attributing knowledge to them.

Notice that this is not the already discussed phenomenon of partial defeat.[58] Justification comes in degrees and a defeater can defeat all of it or part of it. A partial defeater defeats lowers the degree of justification the subject enjoys but without lowering it so much that it is no longer sufficient to support knowledge.

What we have in **The Big Noir** is not a reason that is defeated just enough to put the subject's level of justification right at the threshold for knowledge when it was previously much higher than that. This is not why we are uneasy about attributing knowledge to Marlowe early in the story. Marlowe's reasons give him far less than is needed for knowledge when they are considered as a complete account of what happened. There is simply too much that his IBE doesn't explain. But when we think of those same reasons as a proper part of the explanation that ultimately emerges, they are ultimately good considered in that light.[59]

Something similar happens when a scientist's justification for believing that some event will occur is that their theory predicts it, later investigation reveals defeaters for their theory while yet further investigation reveals that their original theory is a special case of the correct one. It is similar in the following respect. Considered as a complete account, their original theory is not an ultimately good reason for believing the event will take place. Considered as a proper part of the true theory, it is an ultimately good reason. If a Newtonian scientist (for example)

---

[58] See Thune (2010) for more discussion.

[59] It could be objected that when Marlowe reasons this way, there is an implicit "that's all there is to it" assumption or further belief that he has captured that core of the matter. If so, then his reasons are fully defeated (by the rebutting defeater that there is more to it). I respond that his assumption that this is all there is to it is fully defeated but that doesn't mean he lacks coach-class knowledge. Indeed, this case is analogous to cases where a scientist of the past has a theory that is strictly speaking false but correct about some domain they are drawing inferences about. In these cases, I urge the subject's reasons aren't fully defeated even though they think their theory is a complete account of the target domain rather than just a special case of the correct theory. Thanks to Sandy Goldberg for raising this concern.

believes truly that some event will occur because their theory predicts it and their theory is a special case of the true theory of that target domain, is their belief accidentally true? Not exactly, even if the theory on which their prediction is based is thoroughly discredited by evidence they do not possess. The theory is thoroughly discredited as a complete account of physical objects generally because it is unable to explain phenomena within that domain. Nonetheless, there is a sense in which the theory is true and consequently a sense in which the Newtonian's reasons for believing the prediction are ultimately good. Insofar as we restrict our attention to the sense in which the theory is true and the reasons are ultimately good, the Newtonian's belief strikes us as non-accidentally true. If we don't draw the distinction between the two senses in which reasons can be or fail to be ultimately good, then we just feel uneasy saying either that the reasons are or are not ultimately good.

The source of our uneasiness is that Marlowe and the Newtonian have coach-class knowledge. For them to have first-class knowledge, their reasons would have to be ultimately good *simpliciter* rather than ultimately good insofar as this but not insofar as that. Nonetheless, they have a derivative kind of knowledge precisely because there is a derivative sense in which their reasons are ultimately good. Their reasons have a humble form of ultimate goodness because of how they relate to the reasons that are ultimately good *simpliciter*. The proper parthood relation is a little different in each case. In the former, it is just one conjunct of a conjunctive explanation. In the latter, the scientist inadvertently treats a special case of the correct theory as if it were a complete account. The result is that they unwittingly avail themselves of a useful model. In either case though, the subject has a somewhat but not entirely

accidentally true belief because their reasons were a proper part of the reasons that yield first-class knowledge.

Of course, inference to the best explanation is just one form justification might take and the reader might worry that IBE is idiosyncratic in many respects. You might worry that the things I say about it don't generalize to other forms of justification. We will see in sections (II) and (III) that the point generalizes to knowledge that doesn't depend on IBE. What IBE shares with these other kinds of justification is susceptibility to both unqualified and qualified defeat. This is what makes room for the first-class/coach-class distinction.[60]

3.3 Noir Variations

In this section I will consider a variation of the Noir-style case in which the subject's reasons are not a proper part or a special case of the ultimately good reasons but are relevantly similar. I will then apply the lessons learned to the knowledge from falsehoods debate.

Suppose we alter **The Big Noir** so that Marlowe mistook drug residue for counterfeit residue. Suppose further that he comes up with the same account he ends up with at the end of the original case but with no help from Kirk Douglas. That is, he believes that Huston killed Cagney because he learned about the counterfeit operation and the affair. His reason is an inference to the best explanation. He was right, except that it was a drug operation.

Marlowe's reason is ultimately defeated, at least in the strictest sense. Was his belief accidentally true? I feel uncomfortable saying this without qualification. I also feel uncomfortable saying it was non-accidentally true without qualifying my answer. The key is in

[60] Thanks to an anonymous referee for pointing out the need to be clear about this.

the fact that his explanation is defeated *in the strictest sense*. Marlowe's reasons are ultimately defeated in the strictest sense, but they are nonetheless pretty much the same as the ultimately undefeated reasons.[61] Marlowe's explanation resembles the ultimately good one enough for us to want to qualify our statement that his reasons are ultimately defeated. We want to say that he was on the right track, he was basically right or something of the sort. For this reason, the truth of his belief doesn't strike us as a total accident.

Some might ask if Marlowe's justifying reason is really that there is illegal activity of some sort, rather than a belief about the nature of that activity specifically.[62] If so, his justifying reason is not ultimately defeated. Though it is easy enough to imagine a case like this, we are free to stipulate that the case we are dealing with isn't. For whatever reason, let us assume that Marlowe doesn't have a more general belief than his belief about the drug operation or at least that he doesn't base anything on it.[63] If you are still skeptical, consider the following case: I see a cougar and form the belief that I am in danger before I draw any inferences about being confronted by a member of the Puma genus, a large cat, etc. I still know I am in danger. I submit that I have coach-class knowledge if we change the case so that I have a mistaken belief that it is a jaguar from which I infer that I am in danger.

Since I don't form a more general belief about the cat *before* inferring that I am in danger, my belief that I am in danger can't be based on it. Knowledge entails doxastic justification and a reason for believing p can only doxastically justify belief that p if the belief that p is based on it. Since my belief that I am in danger is not based on any mental state having

---

[61] Cf. Klein (2008).
[62] Cf. Ball & Blome-Tillman (2014), Montminy (2014) for two different ways of developing this thought.
[63] Cf. Coffman (2008), Klein (2008).

to do with the size or genus of the cat, my doxastic justification (and consequently knowledge) can't depend on justifying reasons having to do with these things. Granted, I am disposed to believe that there is a large cat as soon as I see it. But a *disposition to believe* can't enter into the basing relation. You can't base a belief off a disposition. A *dispositional belief*, on the other hand, could be the basis for another belief. A dispositional belief requires that I have actually stored the content believed, it is dispositional because I am not entertaining that belief in an occurrent episode, although I am disposed to do so.[64] However, in the case we are imagining I stipulate that I haven't yet processed and stored the content PUMA or LARGE CAT but rather only the more specific content COUGAR. So, I don't have a dispositional belief (as opposed to a disposition to believe) that I am confronted by a large cat or a puma. The burden of proof is on anyone who wants to dispute the psychological plausibility of my stipulation. The burden will be difficult to discharge since it is ultimately an empirical question and work on feature detection, both in AI and mammalian brains, suggests that we first detect low-level features and then proceed bottom-up to increasingly abstract ones.[65]

The Cougar example is just like the noir variation in the relevant respects. The subject forms a justified false belief with a specific content and draws an inference from it before forming a more general true belief that could have, but didn't, appear in his reasoning as a premise. Marlowe's only reason is ultimately defeated, at least in the strictest sense. The question is once again whether he knows. My suggestion is that it is harder to say one way or the other than it is in obvious cases of knowledge or non-knowledge. This is because his reasons are

[64] For more on this distinction, see Audi (1994).
[65] See Gauker (2011), Khaligh-Razavi (2014), Yamins & Di Carlo (2016) & Buckner (2018).

not ultimately good considered in themselves but are ultimately good insofar as they resemble reasons that are ultimately good *simpliciter*. If we just ask whether his reasons are ultimately good or whether he knows without drawing the in-themselves/insofar as they are similar to x-distinction, we will have a hard time giving an unequivocal answer.

I think something along these lines is what is going on in cases of knowledge from falsehood. It was originally thought that one's justification for a true belief depending on a false lemma (that one justifiably believes) was necessary and sufficient for gettierization (Clark, 1963). It was later discovered that one's justification could be defeated without any false lemmas (Lehrer & Paxson, 1968; Goldman, 1976). New evidence might compromise the connection between one's justification and the belief it is supposed to justify without showing the former to consist of any falsehoods. If so, then perhaps false lemmas are sufficient but not necessary for gettierization. However, the sufficiency claim has also been contested. Some cases seem to involve false lemmas in such a way that they don't compromise the subject's knowledge. Consider a case in which a professor believes she has 50 students in her class, so she prints 60 handouts for her lecture just to be safe. She then forms the belief that she has printed enough handouts because she has 60 and she only needs 50. However, she was wrong about having 50 students in her class. She actually has 51. She miscounted.[66] If you worry that she must have also had a more general true belief that there are approximately 50 students because otherwise she would have had no reason to print more than 50 handouts, we can adjust the case so that her TA prints 60 handouts. This way we don't have to explain why she printed 60. When the TA hands them off to her immediately before the lecture, she sees that there are 60 handouts (5 stacks of

[66] This is an adaptation of a case from Warfield (2005).

12, say), reasons that there are 50 students and that 60 is greater than 50. For these reasons (and only these reasons), she infers that there are enough.[67]

Some argue that false lemmas don't always preclude knowledge.[68] Others take the hardline stance that a false lemma always precludes knowledge.[69] I recommend a compromise. The professor's reasons are not ultimately good in the strictest sense, since they depend on her ignorance of the actual number of students in her class. If she weren't ignorant, her justifying reasons would be different, since they are in fact defeated.[70] So, she doesn't have knowledge in the strictest sense. Nonetheless, after remedying her ignorance, her justifying reasons wouldn't have to change much. So, there is some inclination to say that the falsehood is benign, just as there is an inclination to say that she doesn't know in the strictest sense. Why not say both? Why not say that she doesn't know *in the strictest sense,* but she nonetheless enjoys coach-class knowledge on account of her reasons being similar enough to the ultimately good reasons?

It is important to note that the subject's reasons don't just need to be semantically similar to the ultimately good ones.[71] They need to be similar in the right way for the subject to enjoy coach-class knowledge. To bring this point out, consider the following Gettier case. Marlowe believes that Huston is guilty because a generally reliable informant told him that the guilty party was wearing a blue jacket on the day of the crime and the pullman porter told him earlier that

---

[67] Thanks to an anonymous referee for bringing this issue to my attention.
[68]See Saunders & Champawat (1964), Hilpinen (1988), Klein (1996: 106, 2008), Hawthorne (2004: 57), Warfield (2005), Coffman (2008), Fitelson (2010, 2017), Feit & Cullison (2011), Arnold (2013), Hiller (2013), de Almeida (2017), Hawthorne & Rabinowitz (2017), Buford & Cloos (2018), Turri (2019), Luzzi (2019), Zhao (2022).
[69] See Russell (1912: 76), Clark (1963), Armstrong (1973: 198-9), Harman (1973: 47, 120, 1980), Lehrer (1974: 220), Swinburne (2001: 199), Feldman (2002: 36-7), Audi (2003), Lycan (2006: 156-7), Williamson (2007: 145-7), Kripke (2011: 202), Littlejohn (2013), Ball & Blome-Tillman (2014), Schnee (2014), Montminy (2014), Kelp (2016), Borges (2017) and Yong Lee (2021).
[70] Cf. Montminy (2014).
[71] Thanks to Sandy Goldberg for pointing this out.

Huston was wearing a blue jacket on the day of the crime. It turns out the informant just made that up so he could sell Marlowe phony intel. It just so happens the guilty party was wearing a green jacket on the day of the crime. Huston did it and the pullman porter saw him entering the train afterward, but it turns out (unbeknownst to all) the pullman porter is colorblind.

This is a Gettier case rather than a case of coach-class knowledge. The detective believed truly that Huston did it. His reason was that the testimony of the informant and the porter regarding jacket color. However, Huston wasn't wearing a blue jacket. He was wearing a green jacket. That he was wearing a green jacket is an ultimately good reason to think he was the guilty party. Green is similar to blue though, just as 51 is similar to 50. So, why don't we have a case of coach-class knowledge, as we did in the case of the handouts?

Although the detective's reason is semantically similar to an ultimately good one, it is not similar epistemically.

> **Epistemic Similarity:** A good (in the subjective normative sense) reason to believe p, R, is epistemically similar to a good (in the objective normative sense) reason to believe p, N, just in case R and N are semantically similar and the way they support p is also similar.

Reasons are semantically individuated, so semantic similarity matters. But what interests us here is not just their semantic properties. We are also interested in what it is about them that makes them capable of supporting belief that P. Once I apply the epistemic similarity principle to the cases of interest, I will say a bit more about the principle itself and how limited its ambitions are.

Here is how this account of epistemic similarity helps us distinguish the gettiered Marlowe case from the professor with coach-class knowledge. In the former case, the detective's reason for believing Huston did it is the match between two sources of testimony regarding jacket color. The informant told him that the offender was wearing a blue jacket and the pullman porter told him that Huston, one of his suspects, was wearing a blue jacket that same day. These are Marlowe's justifying reasons. Those reasons are factually defeated. The informant was lying and the porter is colorblind. There is a semantically similar undefeated reason: Huston was wearing a green jacket and the culprit was wearing a green jacket. Despite the similarity of green to blue, the undefeated reason doesn't support belief that Huston did it in the same way. Marlowe's justification came from received testimony and the undefeated reasons have nothing to do with that testimony.

In the case of the professor things are different. Her reasons are semantically similar to the ultimately good ones in that 51 is similar to 50. Furthermore, her reasons work in the same way as the ultimately good ones. Both work by positing more handouts than students. This gives the professor a *prima facie* reason to think she has enough. That reason is defeated on account of her being wrong about the exact number. Part of her justifying reason is that there are exactly 50 students in the class, that part of her justifying reason is rebutted by the fact that she missed one student. However, a semantically similar reason that supports the subject's belief in the same way (i.e., by positing a favorable handout/student ratio) is undefeated.

I should note that, in general, similarity is pragmatically squishy and somewhat elusive. I don't expect necessary and sufficient conditions for exactly the right amount of similarity to be forthcoming. Any attempt to use counterfactual analyses in terms of a similarity relation on

possible worlds will be the same in this respect. We can talk about all or nearly all nearby worlds but that still leaves open exactly how many and how nearby they must be. Furthermore, different epistemologists might disagree about the details regarding the way a subject's justifying reasons support a particular belief. My proposal here isn't supposed to help us adjudicate those disputes. Rather, it tells us how to account for grades of knowledge once we have adjudicated them.

Some readers may not feel the pull of the intuition that the professor doesn't know in the strictest sense. Her belief was safe, and this may be what drives the intuition. To make the point even more forcefully, suppose she made 1,000 handouts instead of 60 but was still wrong (by the same margin) about the number of students. Do we still want to say she doesn't know in the strictest sense?

I think we should. Knowledge is non-accidentally true belief. If knowledge is non-accidentally true belief, then somewhat accidentally true belief is somewhat knowledge. We should all agree that the professor's belief is more accidentally true than it would be if she had a true belief about the number of people in the audience. You should grant that even if you endorse some other account of accidentalness, such as the safety account.[72] I will give reasons to prefer my account to the safety account in the final section.

3.4 Post-Gettier Cases

In this section I hope to show that so-called "post-Gettier cases[73]" often fit into the pattern that is the theme of this paper. We should first consider Ginet's Potemkin barn case.[74] S

[72] Cf. Carter (2014) who reaches a similar conclusion by a very different argument.
[73] That is, cases that feel like Gettier cases but involving non-inferential knowledge (so no false lemmas). I get the term from Peter Graham (2000).
[74] From Goldman (1976).

is driving through rural Wisconsin and sees what appears to be a barn on the roadside. Lighting conditions are normal, her vision is good, etc. So, she believes that she is looking at a barn. However, unbeknownst to her, she is in a county that has met with economic hardship. The denizens of this once prosperous county put up barn facades so the economic downturn is less apparent to those passing by. It just so happens that S was looking at the only real barn left in Potemkin barn county.

Does she know it is a barn? The most common answer is "no". It seems that the truth of her belief is accidental, at least to many.[75] Another response is that her perceptual equipment is working properly and her belief was formed on the basis of it, so there is no problem.[76]

Perhaps predictably at this point, I offer a compromise. S enjoys coach-class but not first-class knowledge that she is looking at a barn.[77] To make this move, I must identify a qualified sense in which her reasons are ultimately good, even though they aren't ultimately good *simpliciter*. My thought here is that her foreground reason is ultimately good though it is functioning against background assumptions that are not. Her foreground reason is her perceptual evidence. Perceptual evidence only gives one reason to believe that things are as they appear so long as one is making certain background assumptions, e.g., that one's perceptual system is working properly, lighting conditions are standard, etc. It is epistemically ok for the

---

[75] Some of whom are Goldman (1976), Williamson (2000), Howard-Snyder, Howard-Snyder & Feit (2003), Pritchard (2005, 2012), Lackey (2006, 2009), Reed (2009), Greco (2010, 2012), Luzzi (2010), Kelp (2013), Carter (2014), Goldberg (2015, 2018, 2019).

[76] Some of whom are Brandom (1994), Hetherington (1999), Hawthorne & Gendler (2005), Lycan (2006), Sosa (2007), Turri (2011), Feit & Cullison (2011).

[77] Sosa says the fake barn subject has animal knowledge but not reflective knowledge (2007), which is a different way of trying to split the difference. I have worries about going his way, some of them are about the tenability of the animal/reflective distinction as applied to cognitively mature humans (see Goldberg & Matheson, 2020). Even If those worries can be addressed, I still hope to show that reflective knowledge is graded and account for fake barns in that way.

subject to assume these things. Otherwise, it wouldn't feel like a Gettier case at all but rather an unjustified true belief case. Nonetheless, there is a (factual) defeater for one of her background assumptions. The defeater is that she is in Potemkin barn county where there are many barn facades indiscernible from real barns when viewed from the highway. It is not clear which background assumption is defeated. It does seem clear, however, that she was assuming that conditions weren't as they in fact were. Perceptual evidence gives one justification for believing that things are as they appear, other things equal. This is a clear case in which other things are not equal, though S had no reason to suspect this.

This means that one of her background assumptions has a defeater. This by itself doesn't preclude knowledge. If the defeater has a defeater-defeater, then the defeater could be misleading and her justifying reason could remain intact. There is some inclination to think that this is what is going on here. The fact that she is in Potemkin barn county is a misleading undercutting defeater because she is looking at the only real barn in the county. So, her perceptual equipment is trustworthy in the case in which she finds herself after all. It only seems to be untrustworthy if we consider her perceptual equipment relative to Potemkin barn county in general, but there is a logically stronger reference class relevant here: the particular barn she is staring at. This particular situation is free of the perils with which the rest of Potemkin barn county is fraught. So, the thought is that once we realize which part of Potemkin barn county she is in, the fact that she is in Potemkin barn county no longer gives us reason to doubt her senses. That is, the defeater is defeated in such a way that her original perceptual justification is restored.

I suspect this sort of thinking is part of what (at least tacitly) drives the intuition that she knows. Her perceptual justification is restored, after all. Nonetheless, the truth of her belief still

seems somewhat coincidental. Perhaps this isn't a problem. It is generally agreed that it can be a coincidence that you have the evidence you do without that undermining knowledge.[78] However, the problem is deeper than that. It isn't just that she was lucky to get the evidence she did. Her evidence only has evidential force relative to certain background assumptions. It has evidential force relative to the background assumptions she permissibly makes (i.e., that these are normal conditions) and it has evidential force relative to the background facts of which she is unaware (i.e., this is the only spot in the county where appearances aren't misleading). Her belief is accidentally true because the background assumptions she is actually making are defeated. There are other background assumptions relative to which her perceptual justification would have evidential force, but these are not the ones she is actually making and it wouldn't be permissible for her to make them anyway. I don't have a complete account of what makes it permissible to assume things in the background (though I have some things to say in the next chapter). What seems clear is that despite not explicitly thinking about it, we make some background assumptions and not others at any given time.[79] Furthermore, some background assumptions are epistemically impermissible, at least in the sense that they can't support justified belief or knowledge (e.g., assuming you are in Potemkin barn county under completely normal circumstances). Similarly, others are permissible in the sense that they can support justified belief and knowledge. Furthermore, background assumptions are no less susceptible to defeat than beliefs are. Whatever the correct account of permissible background assumptions turns out to be, it will need to accommodate these things.

[78] Cf. Nozick's (1981) masked bandit case.
[79] Cf. Bach (1985).

The upshot is that insofar as her (foreground) reasons are considered in themselves they are undefeated, and the subject's belief is non-accidentally true. However, when they are considered as part of a package that includes the background assumptions that enable them to acquire epistemic potency, they are defeated because the assumptions on which they depend are defeated.

Fake barns differ from standard Gettier cases in the following way. Consider the Ferrari case from Lehrer (1965). Mr. Nogot is a student in S's class. Mr. Nogot tells S that he (Nogot) owns a Ferrari. S sees Nogot driving it in to campus one day. This perceptual evidence and the testimony from Nogot give S reason to think that Nogot owns a Ferrari. However, unbeknownst to S, Nogot is a pathological liar, and he was borrowing the Ferrari from someone else in S's class. S has no reason to suspect this, so she forms the belief that Nogot has a Ferrari and then existentially generalizes to the belief that someone in her class does, which is true but not knowledge.

S's perceptual evidence is factually defeated by the fact that Nogot was borrowing the car. S's testimonial evidence is defeated by the fact that Nogot is a pathological liar. The perceptual evidence is only perceptual evidence that Nogot owns a Ferrari against the background assumption that people generally own the cars they drive, and this situation is not unusual. The testimonial evidence is only evidence that Nogot owns a Ferrari against the background assumption that testimony about car-ownership is generally reliable and this situation is not unusual. The situation is, however, unusual in such a way that neither background assumption really applies to her situation. The background assumptions that this situation is usual are both factually defeated. In this respect, the case is just like the fake barn case.

However, unlike in the fake barn case, there aren't any facts of which S is unaware that could function as background assumptions relative to which her testimonial and perceptual justification could re-gain traction, were she to learn more about her circumstances. Consider her testimonial justification. Testimony is generally reliable, but Nogot belongs to the smaller and logically stronger reference class of pathological liars. So, his testimony doesn't give one reason to believe what he says. Unlike the fake barn case, Nogot's testimony on this occasion doesn't belong to some yet smaller reference class such that his membership in that class could enable the testimonial justification to re-gain traction. He is a pathological liar, and this is one of the cases where he was indeed lying. This is unlike the fake barn case in which the circumstances in the county are generally misleading, but the subject finds herself in a sub-region of it where they are not misleading.

In the classic Gettier case, there is something wrong with the subject's foreground reasons because there are no (undefeated) background assumptions that a better-informed subject could make and relative to which the same foreground reasons could justify the subject's belief. This differs from the "post-Gettier" fake barn case where the foreground reasons, as such, are fine. They just can't support knowledge when the subject is making the background assumptions she in fact is.

3.5 Objections, Replies & Follow-up Discussion

I will start by addressing a worry people sometimes have when I propose this view. It might appear that first-class knowledge is incredibly difficult to come by if my view is true, as

the world is bound to be rife with defeaters that aren't part of any given subject's possessed evidence.[80]

There are a few things worth saying here. The first is that even if this is true, it wouldn't be that bad. It would only show that real cases of knowledge are rarely paradigmatic. My claim is that we get our grip on what knowledge is by first understanding what it is for a belief to be non-accidentally true. This should be relatively (though not entirely[81]) uncontroversial. As I've said above, I recommend understanding non-accidentally true belief in the way familiar from the defeasible reasoning tradition. I go beyond that tradition in showing how the core ideas of it make room for graded accidentality. So long as it is granted that the subject's believing truly is somewhat accidental and that paradigmatic cases of knowledge are non-accidentally true full stop, the objector grants me all I need. The paradigmatic cases (i.e., first-class knowledge) give us our grip on non-accidentally true belief and, consequently, knowledge. So, they are explanatorily prior to the messier cases of coach-class knowledge. The latter only count as knowledge to the extent that they approximate the paradigmatic cases that give us our grip on the concept of knowledge in the first place.

If the objector is worried about anything, it should be that others in the defeasible reasoning tradition are committed to an implausible form of skepticism. But I mitigate that skepticism by admitting gradations of knowledge. If my view is a kind of skepticism, then those of Klein (1980), Pollock (1986) and Moser (1989) are all more thoroughgoing varieties of

[80] Thanks to Sandy Goldberg and Jennifer Lackey for pressing me on this one.
[81] Knowledge-firsters will of course object. I can't address their concerns here.

skepticism, though the worry about skepticism doesn't tend to arise regarding these views.[82] None of these authors put any constraints on factual defeaters. Since that is the feature of my account that arouses the worry about skepticism, it is just as much a worry about their accounts as mine.

Since it is the lack of constraints on factual defeaters that generates the worry about skepticism, I challenge the objector to come up with a case in which there is a non-restoring factual defeater but nonetheless doesn't involve any degree of accidentality. I conjecture this is impossible because non-restoring factual defeaters are sufficient for Gettierization (i.e., at least the degree of Gettierization one gets in coach-class knowledge).

Some might object along the following lines: Justification is graded and knowledge entails justification, but it doesn't follow that knowledge is graded. Anything above a certain threshold suffices for knowledge. Similarly, one could object that just because accidentality is graded and knowledge is non-accidentally true belief, it doesn't follow that knowledge is graded. Perhaps anything beyond a certain threshold for non-accidentality is knowledge *tout court*. If so, then perhaps there is no first-class/coach-class distinction.

If we go this way, then we will have to get by with two categories (i.e., knowledge and Gettier cases) where I think we need three (i.e., first-class knowledge, coach-class knowledge and Gettier cases). The problem is that the taxonomy is too coarse, and it will require us to group dissimilar phenomena together. If we shrink the taxonomy to two categories, then we either have

[82] Though see Swain (1974) for a preemptive objection to the kind of view shared by these authors calling it "an epistemologist's pipe dream", which sounds like a charge of skepticism to my ears.

to group the cases I have been considering in this paper with Gettier cases or we have to group them with paradigmatic cases of knowledge.

Let us first consider the costs of grouping the cases of interest with Gettier cases. The problem is that they are explanatorily dissimilar. In classic Gettier cases, the subject's reasons are defeated. In the cases of interest here, the subject's reasons are also factually defeated, but they are factually undefeated relative to certain qualifications. Factually undefeated reasons are what separate mere JTB from K. The most natural way of accommodating the point about qualified lack of defeat is that it results in qualified knowledge. In general, if there are jointly sufficient conditions for a property, then meeting one of those conditions relative to certain qualifications will result in having that property relative to those qualifications. Being a male and being unmarried are jointly sufficient for being a bachelor. If you are unmarried, but only relative to certain qualifications, then you are a bachelor relative to those qualifications. If you were the groom in a marriage performed within a religious organization not recognized by the state, then you are a bachelor relative to certain qualifications. We might say you are a "census bachelor". One could insist that knowledge is an exception to this general rule, but that requires argument. You need to tell us what it is about the concept of knowledge that makes it unsusceptible to qualification. Perhaps this can be done, but the burden is on you to show that it can.

Let us now consider the other option: grouping the cases of interest with first-class knowledge. This approach runs into a different version of the same problem. Once again, it requires that we group dissimilar phenomena together. Knowledge is non-accidentally true belief. To understand knowledge, we have to understand what it is for a belief to be non-

accidentally true. We understand non-accidental truth in the strictest sense first. That is, first-class knowledge comes first in the order of explanation. We then come to understand qualified non-accidental truth as an approximation of first-class knowledge. This shows us that first-class and coach-class knowledge need to be distinguished by the theorist.

One could of course grant the claim about the relation of explanatory priority between two types of case but deny that this results in a distinction between grades of knowledge. You might insist that at least some instances of what I am calling "coach-class knowledge" are close enough to the "first-class" end of the spectrum to count as knowledge *simpliciter*.

I grant that coach-class knowledge is quite often close enough *for practical purposes.* Two phenomena might be importantly different for the theorist but similar enough so that we often don't have to distinguish them in practice. My claim is that the epistemologist ought to distinguish the two, even if the layperson can get by without the distinction most of the time. The fact that people typically don't draw this distinction is not evidence that the distinction isn't real or philosophically important. Compare: people used to think Jadeite and Nephrite were the same mineral. Specialists eventually discovered that they were wrong: despite looking the same to the untrained eye, the two differ in molecular structure. I claim that first-class knowledge and coach-class knowledge are like Jadeite and Nephrite in that they are different, but people are generally unaware of the difference. Because people are generally unaware of the difference, it is not lexically marked in ordinary language. It would consequently be a mistake to try to determine whether there is a difference by looking at felicity tests and other linguistic data.[83] In short: we

[83] See Stanley (2005: Chapter 2) for an inquiry into whether knowledge is graded that proceeds by looking into linguistic data.

need some reason to think the explanatory difference between two kinds of case doesn't track a conceptual difference between two kinds of knowledge. The burden is on those opposed to the distinction to show that there is no conceptual difference.

I have argued that either way of trying to collapse the first-class/coach-class distinction comes with a burden of proof. This is not the only reason to retain the distinction. The first-class/coach-class distinction enables us to resolve long standing disputes in epistemology ecumenically. Rather than dismissing one side as simply confused or stubborn, we have excavated insights from both sides of the debates about fake barns and false lemmas. The people who deny that the subjects in such cases have knowledge are right about something important: the difference between (for example) fake barn cases and paradigmatic cases of knowledge. The people who claim that the subjects know are also right about something important: the difference between fake barns and classic Gettier cases. Philosophical work is typically not without difficulties, but it is rarely completely misguided. I have developed a framework for explaining why this is true in a couple of well-known debates in epistemology.

The next thing to consider are the costs and benefits of adopting this view as opposed to a safety[84] account of accidental truth. Safety theorists don't typically talk about degrees of safety. They typically talk about safety outright. This makes sense because they are just interested in whatever amount of it is needed for knowledge. Nonetheless, safety can clearly be graded. Possible worlds where you believe falsely using a similar belief forming process can be more or less remote. If a safety theorist were to try to make sense of graded accidental truth, then the

[84] I use "safety" here broadly enough to include safe-methods accounts such as those of Hawthorne (2004) and Pritchard (2012: 257). In cases where the differences might be relevant, I will discuss them in footnotes.

obvious way to go would be to try to make sense of it in terms of degrees of safety. If safety is what makes for first-class knowledge, then presumably it is a lesser degree of safety that makes for coach-class knowledge.

One point worth briefly noting is that my proposal is compatible with Necessitarianism and the safety view is not. Necessitarianism is the view, endorsed by Spinoza, that there is only one logically possible world. I am not arguing that Necessitarianism is true or even that we should spend much time worrying about it. Nonetheless, it still seems right to say that even if we were to learn (surprisingly) that it is true, we should still want to allow room for the possibility of accidentally true belief. When I imagine myself learning that Spinoza was right all along, I find myself reconsidering many of my philosophical views. But I don't find myself reconsidering whether there are Gettier cases. Yet the safety theorist would have no choice but to reconsider. Nobody with a true belief could have easily believed falsely since there are no nearby worlds at which they or their counterparts could believe falsely. Similarly, I would still want to be able to vindicate the sense that there is an epistemically significant distinction between coach-class knowledge and first-class knowledge even if there are no true counterfactuals and hence no way to draw the distinction in terms of them. This is not because I am sympathetic to Necessitarianism but rather because my commitment to the philosophical significance of these distinctions in epistemology seems orthogonal to my disagreement with Spinoza.

Somewhat relatedly, as was seen in the first chapter, Jennifer Lackey and others[85] have offered Gettier cases such that the subject's belief is safe. Consider the following,

> *Southernmost Barn*: While entering a Midwestern farming community on her cross-country drive, Janice looked at the first barn that she saw, which was on the southernmost end of the field, and formed the corresponding belief 'There is a barn'. As it happens, the barn she saw is the only real one, surrounded by barn façades that members of this community have placed in the field in order to make their town appear prosperous. However, as a matter of strict and unwavering policy, the members of this community always place their only real barn on the southernmost end of their land, since this is where traffic first enters their town. Moreover, thirty years earlier, Janice had lived in a house on the southernmost end of this field in the precise location of the one real barn. Because of her deep interest in her childhood roots combined with the brief period during which she can safely take her eyes off of her driving, she would invariably have looked at only the particular place in the field where the real barn exists. (Lackey, 2006)

The problem is that the belief is made safe in a way that has no bearing on the quality of the subject's reasons. The subject's belief has a defeater, just as in any Gettier case. It is safe despite that defeater not being defeated in a way that restores the epistemic potency of the subject's justifying reason. This shows us that safety can be achieved without eliminating knowledge-undermining luck.

---

[85] Neta & Rohrbaugh (2004), Comesaña (2005), Hiller & Neta (2007), Kelp (2009), Coffman (2010), Bogardus (2014), Goldberg (2015). It matters for my purposes that at least one of these cases works, not that they all do.

If we can make non-knowledge just as safe as knowledge, we can make coach-class knowledge just as safe as first-class knowledge in the same way. We just need to write vignettes with causally relevant facts that don't defeat the defeaters. In some cases, we don't even need to do that. Consider the case of the Newtonian scientist from earlier. They truly predict an event by deriving it from a true description of antecedent conditions and Newtonian laws. They falsely believe that Newtonian physics is a complete physical theory when it is in fact just a special case of a more general theory. I said earlier that this subject has coach-class knowledge but not first-class knowledge. The truth of her belief is more accidental than it would be if she were to know the correct physical theory, but it is less of an accident than a classic Gettier case. But the subject's methods are equally safe in either case. Either way, for them to believe falsely either the laws of physics or the antecedent circumstances would have to be different.[86] By focusing on the quality of the subject's reasons, rather than modal properties that are explanatorily orthogonal to them, I provide an account that focuses solely on the epistemically relevant factors.

Lastly, consider a variation of Warfield's handout cases but where the professor undercounts rather than overcounts. She has 60 handouts and believes that will be enough because she counted 50 students. However, she counted one student twice and she really has 49. Her belief that she has enough is even safer than it would be if she had counted correctly. More would need to change for her to believe falsely that she has enough, since the difference between

[86] Pritchard (2005, 2007) restricts his safety-theory to knowledge of logically and nomically contingent propositions. The proposition known, in my example, is contingent although it is derived from nomic necessities, so it isn't clear that Pritchard's restriction helps here. Furthermore, it doesn't seem that the safe-methods approach he adopts later (Pritchard 2012: 257) meant to deal with necessary truths is going to help here. The methods of the Newtonian and the Einsteinian are equally safe when applied to the domain where their theories are empirically equivalent, and this is the context in which both subjects find themselves.

students and handouts is greater. However, it seems like her belief that she has enough is epistemically worse off rather than better.

3.6 Conclusion

I hope to have shown that knowledge admits of gradations, not because justification does, but because accidental truth does. There are several ways this can happen. I discuss a different one in each section and there are likely to be others I haven't considered. This is significant in part because it allows us to bring the debates about knowledge from falsehood and fake barns to ecumenical resolutions. Furthermore, I hope to have shown that the defeasible reasoning tradition can give a better account of what is going on in these cases than a modal account.

So far, I have defended and further developed the defeasible reasoning tradition. The task of the next two chapters is to connect the epistemological theory that results with the philosophy of mind and the philosophy of cognitive science.

Chapter 4: Defeasible Reasoning & Cognitive Science

"As I understand it, being justified is a status one has in relation to a person or group of people. If that's not what epistemologists have in mind when they speak of epistemic justification, I have no
idea what they mean."

-Sydney Morgenbesser

Epistemologists who want their work to be continuous with and relevant to work done in Cognitive Science tend to be externalists about justification. The most obvious way of forging the connection between Epistemology and Cognitive Science is to reduce epistemic vocabulary to the kind of non-epistemic vocabulary that employed by empirical researchers. For example, one might either reduce or at least specify a supervenience base for terms such as "justification", "epistemic reason", "knowledge", etc. in terms such as "belief-forming process", "reliability", etc. I have already argued against such an approach in Chapter 1. However, that leaves the question of how Epistemology as I conceive of it relates to the conception of ourselves and our place in the world that emerges from Cognitive Science. The purpose of this chapter is to answer that question.

I will proceed by showing that work on social intentionality[87] in cognitive science and work on defeasible reasoning[88] are mutually illuminating and mutually supporting. The work in cognitive science helps us understand why justification and knowledge, as understood by proponents of the defeasible reasoning tradition, matter for creatures like us. The work in

[87] The leading proponent of this movement is Michael Tomasello (1999; 2014; 2021). Related ideas are pursued in Karmiloff-Smith (1992), Billig (1996), Dunbar (1998), Frawley (1997), Dessalles (2007), Mercier & Sperber (2011; 2017), Godfrey-Smith & Yegnashankaran (2011), Scott-Philips (2015).

[88] Lehrer (1965; 1970; 1974), Lehrer & Paxson (1968), Hilpinen (1971), Swain (1974), Barker (1976), Klein (1971; 1976; 1980), Pollock (1986), Moser (1989), Schroeder (2015), de Almeida & Fett (2015). See Shope (1983) for a useful overview of the history of defeasible reasoning in epistemology.

defeasible reasoning helps us understand how the game of giving and asking for reasons does what the cognitive scientists say it does. I will develop an account of the foundations of knowledge that draws equally from both traditions to this end.

I begin by motivating the claim that thought differs from mere information-processing because it involves cognitive self-regulation on the part of the subject.[89] I will consider three kinds of self-regulation and three corresponding kinds of thought associated with them.[90] After describing the most basic kind, I generate the second kind by describing a way that the capacities at play in the previous kind of self-regulation can take a social turn by interiorizing a form of social cooperation: second-personal cooperation (more on "interiorization" in §2). I then generate a third kind that involves the capacities at play in the previous one taking another turn by interiorizing a more complicated form of social cooperation: third-personal cooperation. The epistemic component of this form of social cooperation is the game of giving and asking for reasons. The defeasible reasoning tradition gives us the normative kinematics of this game. A capacity for higher-order reflection will be shown to be necessary for the third kind of thought. Reflection's philosophical significance consists, then, not only in the benefits of the highly adaptive form of social cooperation it makes possible. It also makes our thought different in kind from other forms of thought.[91]

In section (I) I begin my taxonomy of kinds of thought or, as Tomasello (2014) puts it, "intentionality". Each kind of intentionality emerges from the previous kind, both evolutionarily and conceptually. In section (I) I do some stage setting and discuss the most basic kind,

[89] Here I will be developing ideas found in Hurley (2003) and Camp (2009).
[90] Here I follow the taxonomy found in Tomasello (2014).
[91] This aspect of its significance is often overlooked by reflection's detractors. See for example Kornblith (2012).

"Individual Intentionality". In section (II), I discuss the emergence of "joint intentionality" when self-regulation takes a social turn by interiorizing a form of social cooperation. In section (III), I discuss the emergence of "collective intentionality" when self-regulation takes another turn by interiorizing a more complicated form of social cooperation. Throughout section (III) we will encounter claims made in the cognitive sciences that seem right but cry out for further explanation. In section (IV) I lay out the core claims of the defeasible reasoning tradition. In section (V) I draw on the defeasible reasoning tradition to further explain the claims from section (III). Part of the explanation consists in my account of the foundations of knowledge that ties the work on social intentionality together with the work on defeasible reasoning.

4.1 Background on Cognition and Individual Intentionality

Thought isn't something that simply occurs within us: it involves cognitive activity on our part. This is what differentiates genuine thought from varieties of cognition that are less robust, such as stimulus-response learning and the "innate releasing mechanisms" of early ethology.[92]Unlike thought, these are sub-agential processes. They occur within the agent, but the agent takes no active role in them.[93]

A common narrative in the history of cognitive science is that the discipline was born with the overthrow of behaviorism. This happened when mentalistic vocabulary was no longer used as shorthand for behavioral dispositions but rather to refer to theoretical entities (often agent-level cognitive states) capable of explaining those behavioral dispositions.[94]Part of the

[92] Cf. Lorenz (1965), Tinbergen (1951).
[93] By "agent", I don't necessarily mean "moral agent". I consider three kinds of agency below, only the last of which is moral agency.
[94] See Bruner (1972), Fodor (1975), Baars (1986), Miller (2003). Greenwood (1999) offers criticism but agrees with something close to this.

reason the overthrow was needed in the first place was the difficulty of accounting for flexible and creative problem solving, sometimes on the first try, with operant and classical conditioning alone.[95] It became increasingly apparent that we can't explain the cognitive feats of humans and many non-human animals without positing an agent-level ability to monitor the processing, storage and deployment of information.[96] This reflects the fact that these creatures think and the capacity for thought requires a cognitively active agent.[97] If you take issue with my use of "thought" for whatever reason, I'll give you the term. You will presumably acknowledge more and less robust kinds of cognition. Read my remarks about thought as being about an explanatorily significant kind of cognition, more robust than other kinds due to the activity of the subject.

Going forward, I will flesh this picture out by first identifying the basic capacities that make a subject cognitively active. This will enable us to distinguish different forms of thought later by identifying qualitative changes in these capacities.

The first form of thought I will consider is "Individual Intentionality".[98] Individual intentionality involves a capacity for causal inference. A capacity for causal inference is required because robust cognition requires "some kind of mental representation that goes 'beyond the information given' to direct perception" (Tomasello & Call 1997: 8). Similarly, Elisabeth Camp argues persuasively that the capacity for stimulus-independent representation is an important aspect of cognitive activity (Camp 2009: §3). To enjoy individual intentionality, the subject must

[95] Camp (2009) makes a similar point.
[96] In the case of non-human animals, pioneering work includes that of Griffin (1978), Premack & Woodruff (1978), and Cheney & Seyfarth (1980).
[97] Cf. Hurley (2003), Camp (2009), Tomasello (2014; 2021).
[98] I get the term from Tomasello (2014: Chapter 1).

be able to represent situations both as goal states and as already-obtaining states and flexibly adapt her behavior to discrepancies between the two. Activity is a kind of control and, as Susan Hurley puts it, "Control just is the maintenance of a target value by *endogenous* adjustments for *exogenous* disturbances" (Hurley 2003: 235, her italics). This involves a model of the causal structure of the relevant features of the environment and the outcomes that would result from various interventions in it.

Plausibly, this involves a capacity to simulate the outcome of possible courses of action offline.[99] The subject considers a possible course of action directed at the goal state, "presses play" and imagines the outcome of that action as determined by her causal model of the situation. As Barsalou puts it, "the ease of simulating a scenario underlies the acceptability of a causal explanation" (1999: 606). This enables the subject to anticipate the results of courses of action she has never performed before, which makes individual intentionality more robust and potentially more useful than mere operant conditioning. For example, in an experiment performed by Manrique et al. (2010), chimpanzees were presented with a novel food extraction problem. Its solution required a tool of certain dimensions in a different room out of sight. Chimpanzees could sometimes solve the problem on their first trial, demonstrating their possession of a causal model they retained even after they went to the other room and their proximal stimulation changed. As Tomasello explains, "They then simulated the use of at least some of the tools and the likely outcome in each case" (2014: 16).

[99] See Barsalou (1999; 2005; 2008).

Another important difference between individual intentionality and mere associative learning is the capacity for self-monitoring that it involves. The ability to simulate the outcomes of possible courses of action involves the subject being sensitive to its own simulations as well as the external features of the environment that it is tracking in the simulation. The subject needn't conceptualize the simulation as a simulation but must be implicitly sensitive to the fact that it is one. This involves a capacity for uncertainty monitoring: the subject must be able to determine how well equipped they are to simulate the situation accurately offline. Along with causal inference, this capacity comprises the most basic form of executive functioning, and it is what makes subjects with robust individual intentionality capable of "taking charge" in a way that creatures with only stimulus-response connections and fixed action patterns are not. So long as there is a capacity to reconsider one's own mental states adaptively and implicitly, the creature is engaged in cognitive self-monitoring. There is empirical evidence that dolphins, pigeons, rats, monkeys and apes enjoy this capacity (Cf. Smith, Couchman & Beran 2014).[100]

Many creatures with individual cognition have a primitive theory of mind as well (cf. Premack & Woodruff 1978; Tomasello & Call 2008). That is, they encode information about the intentional states of others and factor that into their decision-making. Creatures with individual cognition are sometimes social animals, although they understand their relation to their conspecifics in a "Machiavellian" way (cf. Humphrey 1976). That is, conspecifics are much like the other features of their environment with which they must cope, except they have intentional states.[101]

---

[100] See also Hampton (2001), Washburn, Smith & Shields (2006), Kornell, Son & Terrance (2007), Call (2010).
[101] See Hare et al. (2000) for more on this.

Here is a more robust form of social thought, one I will call "joint intentionality".[102] Instead of understanding one another as agents going our separate (though sometimes mutually beneficial) ways, we might understand ourselves as jointly pursuing a common goal and each performing complementary roles to that end.

This requires a capacity for socially recursive inference. That is, in order for us to form a joint intention to pursue a common goal, we have to each understand that the other understands that we understand that… this is what we are doing. The recursive reasoning will only ever be taken a few layers deep in occurrent mental episodes, but the creature implicitly understands that the pattern of inference can go on in the same way indefinitely (Tomasello 2014: 38).[103][104]

Socially recursive inference makes "common ground"[105] between the parties to the joint activity possible. The common ground will be formed by *ad hoc* gestural communication and pantomiming (Tomasello 2014: 59ff.) since discursive reasoning and communication only emerges alongside the next kind of intentionality to be considered.

This kind of common ground makes possible another key feature of joint intentionality: social self-monitoring. Since we each have to know a certain amount about each other's mental states to form a joint intention in the first place, when we have a joint intention, we will also be able to understand what we are doing from the perspective of our counterpart. If you and I are foraging together, I know enough about your mental states to assess my own performance from

---

[102] Again, in line with the terminology of Tomasello (2014: Chapter 3).
[103] Cf. Clark (1996), Tomasello (2008).
[104] So far as I can tell, the first three layers are the important ones. Socially recursive reasoning is important because it enables us to monitor ourselves from the perspective of someone else. That requires three layers. So, although I use the term "recursive", it would be ok for my purposes if the competence itself were limited to three iterations.
[105] This is somewhat like Stalnaker's (1978) account of common ground, save that it is "quasi-propositional". For more on this, see Tomasello (2008; 2014: Chapter 3).

your perspective. So, even though it might be tempting for me to free ride and let you do all the hard work, I am also capable of understanding the situation from your perspective. Joint intentionality does not include the ability to make this understanding explicit. That would require discursive capacities foreign to joint intentionality. Creatures with joint intentionality can only understand this implicitly via simulated assessment from the perspective of their collaborator.

To sum up, second-personal activity involves each participant playing a distinct role in a shared activity. When this kind of activity is interiorized, the subject incorporates both roles into her own thinking by toggling between viewing the activity from her own perspective and the perspective of the agent fulfilling the complementary role. This involves evaluating her own performance from the perspective of the other agent through simulated self-assessment. In this way, she views herself second-personally, evaluates herself from the second-person perspective and regulates her behavior in light of this evaluation. This is what "interiorization" involves.

At this stage we have seen a purely first-personal form of thought and a second-personal form of thought emerge from it. This latter form of thought is distinctive in that it involves interiorizing the perspective of a collaborator into one's own thinking and using that perspective to assess prospective courses of action and regulate them in a way tailored to cooperative activity. The final turn is the emergence of third-personal thought or "collective intentionality (cf. Tomasello 2014: chapter 4). Here we build on the capacities at play in second-personal "joint intentionality" but with a twist: the subject interiorizes the perspective of the "generalized" other rather than just the perspective of this or that collaborator. This enables one to view oneself third-personally and evaluate oneself from the third-person perspective.

As in the case of joint intentionality, decision-making is collaborative. However, now the coordination is discursive. At this stage, coordination goes beyond pantomiming and involves giving and asking for reasons. Instead of gesturing at the antelope tracks and hoping you will recognize what I think we should do, I can now explicitly state the intention I recommend we form together and explicitly support my recommendation with reasons you could either accept or challenge. Indeed, the primary biological function and philosophical significance of reasoning is not the regulation of individual thought, but rather interpersonal argumentation. The role reasoning plays in regulating the private thought of individuals is a derivative phenomenon that results from the interiorization of interpersonal argumentation. As Mercier & Sperber (2011: 60) put it,

> Reasoning contributes to the effectiveness and reliability of communication by allowing communicators to argue for their claim and by allowing addressees to assess these arguments. It thus increases both in quantity and in epistemic quality the information humans are able to share. Claiming as we do that this role of reasoning in social interaction is its main function fits well with much current work stressing the role of sociality in the unique cognitive capacities of humans.[106]

By "reliability" they mean "truth conduciveness". So, one of their claims is,

(1) The game of giving and asking for reasons is truth conducive.

[106] They then go on to cite the following, Byrne & Whiten (1988), Dunbar (1998) Dunbar & Shultz (2003) Hrdy (2009) Humphrey (1976), Tomasello et al. (2005), Whiten & Byrne (1997).

I will refer to the activity of engaging in interpersonal argument as "the game of giving and asking for reasons" following Sellars (1956/1997). I do this because, as I will argue later, we can best understand interpersonal argumentation epistemologically by drawing on the defeasible reasoning tradition and defeasible reasoning, I hope to show, is helpfully thought of as a kind of game (cf. Lehrer 2017).

Returning to Mercier & Sperber, the social emphasis of their account sometimes makes people think that truth-conduciveness is unimportant on Mercier & Sperber's view (e.g., Tomasello 2014: 110). Some of their remarks suggest this. For example, at one point they say that rather than the improvement of knowledge being the function of reasoning in humans, it is actually persuasion (Mercier & Sperber 2011: 57).

These remarks are misleading. On their account, the purpose of inference generally is the correction and augmentation of information available to a cognitive system (Mercier & Sperber 2011: 58). Reasoning consists in inferences about arguments (Mercier & Sperber 2011: 58-9). Although these inferences are about arguments, the arguments themselves can be about anything (Ibid). So, the inferences indirectly augment and correct information about whatever the arguments concern. The fact that this latter information is corrected and augmented indirectly does not mean it is corrected and augmented accidentally. If inferences about arguments couldn't give us information about things other than the arguments themselves, argumentation would be a self-contained game with no adaptive value.

In light of all this, we should not take Mercier & Sperber to be saying that persuasion rather than truth-conduciveness is the function of reason. The most charitable interpretation is

that the function of reason is to be truth-conducive in argumentative contexts specifically.[107] This is their explanation of why reasoning is poor (i.e., not truth-conducive) in non-argumentative contexts, such as the Wason selection task (Mercier & Sperber 2011: 64).[108]

(1) sounds plausible, but empirical researchers haven't, to my knowledge, offered a deeper explanation of it. This is fair enough since it is a plausible claim, and their primary task is to explain other things in terms of it. Nonetheless, it would be good to have a deeper explanation of how the game of giving and asking for reasons is conducive to true belief. In fact, the connection between the game of giving and asking for reasons and the truth often takes a backseat to its function in facilitating social coordination in the empirical work. Indeed, when Tomasello talks about good reasons he sometimes puts "good" in scare-quotes.[109] The thought seems to be that the reasons can only be "good" in a sociologically relativized way constructed by the attitudes of the participants in the game. While Tomasello is right that the game of giving and asking for reasons is a social phenomenon, we should be hesitant to understand its norms in a way that severs the connection between the beliefs we offer reasons for and the truth of those beliefs. When this connection is severed, it is hard to see how the "reasons" put forth within the game are reasons to believe anything at all. It also makes (1) seem surprising. Since we want to explain (1) rather than have it be an anomaly on our theory, we should pursue other possibilities.

[107] See their (2011) exchange with Godfrey-Smith & Yegnashankaran (2011) for further support of this interpretation.
[108] Cf. Resnick et al. 1993, Blum-Kulka et al. 2002; Neuman 2003; Weinstock et al. 2004.
[109] For example, see Tomasello (2021: 20).

Going forward, I won't commit myself to any claims about what grounds the norms of the game or what the ultimate source of their normativity is.[110] Rather, I will show that we can draw on the defeasible reasoning tradition to articulate the normative kinematics of the game and doing so will help us flesh out the picture Tomasello and others give us in a way that helps us answer some of the outstanding questions their work leaves us.

One of Mercier & Sperber's main reasons for claiming (1) is that it helps explain how reasoning can be adaptively beneficial for humans even though we are prone to fallacious reasoning in non-argumentative situations. People display poor logical reasoning skills in contexts detached from interpersonal argument (Evans 2002: 981) but do much better when not so detached (Resnick et al 1993; Pennington & Hastie 1993: 155; Blum-Kulka et al 2002; Thompson et al 2005; Hagler & Brem 2008).[111]

The primacy of social reasoning does not mean that solitary reasoning is dispensable nor is it to deny that it also serves the function of improving our cognition. Rather, the point is that the quality of individual reasoning is improved as the reasoner's inner monologue starts to sound more like a dialogue (cf. Mercier & Sperber 2011: 73[112]). For instance, Descartes' *Meditations* are the monologue of a solitary meditator, but that meditator has a knack for anticipating points at which others will resist and uses those points of resistance as a way of improving the

[110] Tomasello draws heavily on Searle (1995; 2001) in his account of where the norms come from. I remain agnostic.
[111] Many of the studies cited here test the ability to recognize formal fallacies. Similar findings pertain to informal fallacies. See Hahn & Oaksford (2007), Neuman (2003), Neuman et al. (2006), Weinstock et al. (2004), Rips (2002).
[112] Baier (1981: 182) makes a similar claim about the methodology of the *Meditations.* There she also sympathetically discusses a view of human thought much like the one I propose here and its relation to Descartes' views.

arguments under consideration.[113] We might say that the monologue reads like a dialogue for this reason. My claim here is that individual thought serves its epistemic-regulative function by internalizing interpersonal argumentative dialogue, the purpose of which is social coordination on truth, and this is why individual thought that better mimics interpersonal dialogue better serves its regulative function (cf. Godfrey-Smith & Yegnashankaran 2011: 80). It may also explain why certain congenital disorders that disrupt inner speech (e.g., autism and Williams Syndrome) also disrupt cognitive self-regulation (cf. Frawley 1997).[114]

This is related to another point Mercier & Sperber make in the above quote: reasoning contributes to the effectiveness of communication. "Effectiveness" can mean many things, but they seem to be using it synonymously with "persuasiveness" here. This is particularly important because social cooperation obviously requires a fair amount of doxastic coordination. However, it would be good to have a deeper understanding of how interpersonal reasoning contributes to coordination and, most importantly, how to square this with the truth-directedness of doxastic states. So, I will now mark another important claim to be revisited and further explained in section (V),

(2) The game of giving and asking for reasons is coordination conducive.

It is important to bear in mind here that it is not just our social relations with one another that change as we shift from joint to collective intentionality. As with the shift from individual to

[113] Consider, for example, the transition from the sensory fallibility argument (7:18) to the dream argument (7:19) in Descartes (2009).
[114] The above gives us reason to think that the epistemological significance of inner speech goes deeper than is typically thought. For alternative accounts of its significance, see Jackendoff (1996), Bermudez (2003) and Munroe (2021).

joint intentionality, a new form of self-regulation emerges. In joint intentionality, we saw a new form of self-regulation that involves self-monitoring and self-assessment from the perspective of this or that potential collaborator. In collective intentionality, a new form of self-regulation emerges when the standards of interpersonal argumentation are interiorized. That is, the same norms of argumentation that we can use to coordinate with others can also be used to regulate our own thought by simulating self-assessment from the standpoint of an imagined audience. However, it needn't be from the standpoint of anyone in particular. Collective intentionality was selected for in virtue of its ability to foster large-scale collaboration (Tomasello 2014: 80ff.). To this end, it enables us to coordinate with others generally, even those with whom we have yet to jointly attend to local goings on. As Tomasello puts it,

> The Capstone of all this- recognized by modern thinkers who take a sociocultural view of human thinking- is the internalization of these various interpersonal processes of making things explicit into individual rational thinking or reasoning. Making things explicit to facilitate the comprehension of a recipient leads the communicator to simulate, before actually producing an utterance, how his planned communicative act might be comprehended-perhaps in a kind of inner dialogue. Making things explicit to persuade someone in an argument leads the disputant to simulate ahead of time how a potential opponent might counter his argument. (Tomasello 2014: 112).

The philosophically significant claim here is that individual thought results from the interiorization of paradigmatically interpersonal activities, particularly those involving persuasion. In other words,

(3) For those enjoying collective intentionality, private individual thought is the interiorized form of the game of giving and asking for reasons.

This claim will be important later. However, that is not because I will draw on work in epistemology to give a deeper explanation of it but rather because I will draw upon it to give a deeper explanation of the significance of work done in epistemology.

With the internalization of interpersonal argumentation, we see the emergence of a new, essentially reflective, form of self-governance. We interiorize the standpoint of the critic (not this or that critic, but rather the moves available to any critic whomever they might happen to be) and use that interiorized standpoint to regulate our own cognition.

One might worry that although this kind of self-conscious and deliberate rehearsal before an imagined audience is something we do sometimes, it isn't something we do all the time. This is true, but it is consistent with my proposal. Since self-regulation is a necessary condition for robust cognition, the kind of self-regulation we enjoy determines the kind of cognition we have, and the epistemic standards appropriate to assessing intentional states of that kind. Even those states the subject doesn't bother to regulate can be epistemically assessed by reference to the standard. Since self-regulation here takes the form of interiorized argumentation, the standards of argumentation are the standards appropriate to its assessment. I will have more to say about these standards in the next section. To foreshadow a bit, the normative kinematics of the game of giving and asking for reasons have a default and challenge structure. Since the subject is default entitled to rely on the most often-used belief-forming processes, we shouldn't expect the need for constant simulated self-assessments. Similarly, we shouldn't expect most claims made in interpersonal communication to be challenged.

So far, we have been considering the evolutionary function of the game of giving and asking for reasons and the way higher-order epistemic reflection emerges from its interiorization. The evolutionary function is agonistic collaboration meant to coordinate on the truth. None of this is to deny that people sometimes play the game of giving and asking for reasons in bad faith or that the outcome of argumentative exchanges can be influenced by social power relations in untoward ways. Rather, it is to understand the practice in its ideal form, which will then allow us to see these things as corruptions of it. It is only by understanding the practice in its healthy form that we can later recognize pathology in it.

To better understand the ideal form of the practice, we need to consider what makes it capable of serving the purpose of truth-conducive generalized belief coordination. I will now draw on the defeasible reasoning tradition in epistemology to make sense of this.

4.2 Defeasible Reasoning and Collective Intentionality

(1)-(3) require further explanation and defense. To provide that explanation and defense, we need to turn to the defeasible reasoning tradition in epistemology. As I've argued in previous chapters, the key insight of the defeasible reasoning tradition is that justification and knowledge are a matter of the standing of the subject's reasons.[115] To determine whether the subject has the former, we must look at those reasons from the subject's perspective. To determine the latter, we look at them *sub specie aeternitatis*. Either way, the procedure for epistemic assessment is the same, it is just performed relative to a different body of potential defeaters.

[115] Some working in the tradition, such as Peter Klein (2017), say it is a matter of the quality of her evidence. I talk of reasons to better connect the theory of justification with the philosophy of mind developed earlier in the paper.

I propose we understand the game of giving and asking for reasons in terms of the way that reasons put forward within that game are epistemically assessed. I claim that the defeasible reasoning tradition gives us an account of this at a structural level. That is, according to the defeasible reasoning tradition, epistemic statuses are a function of rational relations between the subject's justifying reason(s) and other potentially countervailing reasons.

We can think of it like a card game. When it is time for me to put my cards on the table, we assess my hand by comparing it to the other hands on the table. For me to have any shot at winning, my hand must at least be a provisional winner (i.e., must involve a *prima facie* reason). A provisionally winning hand is an ultimately winning hand just in case either a) no other hand on the table hand even provisionally overrules it (i.e., no defeaters whatsoever) or b) somebody else's hand provisionally overrules it, but only provisionally. In the case of b), your hand is provisionally overruled, but the hand that overrules the hand that overrules your hand does so in such a way that your hand goes back to the top of the leaderboard, rather than a different hand speaking in favor of the same propositional content. This is the upshot of the restoring/non-restoring point made earlier about defeaters.

4.3 Explaining (1)-(3)

Here are (1)-(3) again, but this time with a bit more flesh on (1) and (2) meant to foreshadow the explanation promised earlier,

1a) The game of giving and asking for reasons is truth-conducive because it is potentially self-correcting.

2a) The game of giving and asking for reasons is a coordination-conducive procedure: we lay down our cards and defer to the winning hand.

3) For those enjoying collective intentionality, private individual thought is the interiorized form of the game of giving and asking for reasons.

The game of giving and asking for reasons works with one party giving a *prima facie* reason to believe something and someone (usually someone else) assessing the reason by checking for defeaters. If there is none, then the *prima facie* reason is an all-things considered reason. If there is a defeater, then we need to determine if that defeater is itself defeated. If it is not defeated, then the subject's *prima facie* reason is not an *ultima facie* reason. If it is defeated, then the *prima facie* reason is an *ultima facie* reason only if the defeater-defeater is restoring. This tells us what it is for a reason to be an *ultima facie* reason relative to a body of evidence. The next thing to determine is which body of evidence constitutes the ideal endpoint of GoGAR.

I propose that the ideal end point of GoGAR is a body of total evidence that is coextensive with the facts. In other words, the ideal end point of GoGAR is a stage where the distinction between knowledge and justification collapses. At this stage, it is impossible to have a justified false belief or a Gettiered belief. If you have a false belief, then your justification for it is defeated by some fact. For this reason, GoGAR is maximally truth conducive at its ideal end point. That is, if you have an undefeated reason to believe p at the end point of GoGAR, then p is true.

Of course, no finite subject finds herself at the ideal end point. What now needs to be shown is that finite subjects playing GoGAR can advance themselves closer to the ideal end point by playing the game. Since the ideal end point is maximally truth conducive, GoGAR becomes increasingly truth conducive as we approach it. The ideal is regulative since finite

subjects can never reach it. We can nonetheless improve ourselves epistemically as we approach it.

To understand how GoGAR can advance participants toward the ideal end point, we must understand its self-correcting potential. Although actual inquiry is not fully ideal, we will best understand the self-correcting potential of GoGAR by considering the case of ideal inquiry in which this potential is fully realized. We will then be in a position to look into the extent to which that potential is actually realized.

To understand the self-correcting potential of GoGAR, we need to understand why there is a need for self-correction in the first place. The need for self-correction arises because it is methodologically necessary that we take certain things for granted in inquiry, but to do so without dogma we must take those things for granted only provisionally. It is methodologically necessary to assume that tokens of certain belief-types are for the most part true because we would otherwise face a regress. These tokens are our foundational beliefs.

Since the assumption that they are mostly true is methodologically necessary, we inquire into it through conducting inquiry that provisionally assumes it. That is, we assume that our foundational beliefs are mostly true but expect that further inquiry within the scope of that assumption will lend further confirmation. The assumption is only provisional because further inquiry could do the opposite. Recall that, according to the defeasible reasoning tradition, foundational beliefs don't depend positively on further reasons for support but depend negatively on the absence of defeaters. That is, a token perceptual belief does not depend for its justification on positive reasons to think perception is reliable in this case, but rather depends negatively on

the absence of defeaters giving one reason to think it is unreliable in this case. In Chapter 3 I argued that the same applies to our background assumptions in forming particular (for example) perceptual beliefs. Here I propose that the same applies to the background assumptions of our epistemic practices. The background assumption that perceptual beliefs in general are mostly true requires no positive support but nonetheless depends on the absence of defeat.

Furthermore, part of the inquiry conducted within the scope of these background assumptions faces the burden of explaining why each kind of foundational belief is mostly true. For example, we must assume our perceptual states are mostly accurate to do perceptual psychology. Perceptual psychology is an empirical science that requires testing hypotheses by observation. However, we expect our findings in perceptual psychology to further vindicate the background assumption by accounting for the mechanisms that enable us to discriminate objects perceptually.[116] Complacent inquiry that doesn't probe into its own background assumptions in this way is conceivable, but not without epistemic cost. The cost is that the foundations become a fixed tradition rather than a provisional starting point. The practice of giving and asking for reasons itself becomes epistemically defective when its practitioners don't probe the background assumptions of the practice. This is because they introduce more dogma than is methodologically necessary and consequently impede the self-correcting potential of the game of giving and asking for reasons.

It is always possible that inquiry within the scope of these background assumptions will fail to provide the sought-after vindication. In such a case, the strength of defeat for the

[116] This is exactly what we see, for example, in when Marr (1982) explains the computational processes that enable us to discern luminance contours.

background assumption is proportional to the maturity of the discipline investigating it. So, although the background assumptions don't require positive support (only the absence of defeat) for their epistemic potency, at a certain point in inquiry the absence of positive support will itself be a defeater. This is because, as the discipline investigating the assumption becomes increasingly mature, it becomes increasingly reasonable to think that if the assumption were true, the discipline would have been able to explain why by now. This fact makes the provisional reliance on the background assumptions less dogmatic than it might at first appear.

The picture I recommend is one according to which it is methodologically necessary that we assign some non-inferentially formed beliefs foundational status. Which ones receive this status is determined by our intellectual inheritance. Our cultural initiation involves the inheritance not only of beliefs[117] but also a practice of argumentatively defending them (cf. Tomasello 2014: 111). Part of that practice will involve treating tokens of some belief types as the unmoved movers of the practice. This involves the assumption that those beliefs are for the most part true. This assumption is, of methodological necessity, brute and uncritical at first (cf. Alfano & Levy 2019). However, it is not mere dogma because the inquiry conducted within its scope has the potential to put that very assumption in jeopardy.

Let us now consider ideal inquiry for agents with collective intentionality. We have seen that the need for background assumptions that tokens of certain belief-types are mostly true. Let us consider a case in which these assumptions are true for the agents in question. Furthermore,

[117] See Clark (1996) for more on this.

let us imagine a community such that nobody in it suffers from memory loss. This is obviously unrealistic, but it is nonetheless a useful simplification that we will abandon later.

We have a community of agents with collective intentionality conducting inquiry by playing GoGAR with one another. They are engaged in a form of agonistic collaboration: they are jointly pursuing the truth by exchanging information and occasionally debating with one another about which claims are true. Any time one of them forms a belief, she informs the others. They check for defeaters and, if there are none or the defeaters are defeated, they adopt that belief. Since their foundational beliefs are, we are assuming, mostly true, the web of beliefs for any given one of them is increasing in number as time goes on. Defeasible reasoning is not only an organon, but also a canon. So, sometimes they acquire defeaters for beliefs that were previously formed and their web of beliefs momentarily contracts. However, expansion exceeds contraction since the foundational beliefs are mostly true, the defeasible inferences drawn from them are as well and some of the beliefs that are revised when a defeater is discovered are re-instated later when a defeater-defeater is discovered.

As the web of beliefs expands and the false ones are gradually uncovered and discarded by the critical resources of defeasible reasoning, the web gradually approaches the web that constitutes the ideal end point of GoGAR: the web that includes belief in all and only facts.

We now see that GoGAR is *potentially* self-correcting in a way that makes it increasingly truth conducive as inquiry proceeds. We now need to ask whether it is *actually* truth conducive for us. There are clearly many respects in which actual inquiry does not resemble ideal inquiry. We suffer memory loss. We don’t share all the information we acquire. Ideal inquiry takes the

form of agonistic collaboration, but non-epistemic motives curtail earnest collaboration frequently. In ideal inquiry, the background assumption that foundational beliefs are mostly true is itself true. The extent to which actual human inquiry approximates ideal inquiry in this respect is debatable.

I won't try to determine the exact effect any of these factors has individually on the relation between actual and ideal inquiry. Rather, I will argue that despite these impediments, human inquiry approximates ideal inquiry enough to serve the needs of cumulative cultural evolution. Cumulative cultural evolution, of the sort exhibited by behaviorally modern *homo sapiens* at any rate, requires incremental epistemic improvement. This requires that we gradually approach the ideal end point of GoGAR, at least in some domains.

Tomasello (1999) persuasively argues that cumulative cultural evolution in general is made possible by inheriting a goal-directed practice from one's ancestors.[118] Without this inheritance, everyone would need to reinvent the wheel. Once one has inherited the goal-directed practice, one can refine its methods so that they are more conducive to the goal. This makes intergenerational incremental improvement possible.

The game of giving and asking for reasons facilitates the epistemic component of cumulative cultural evolution. As Kim Sterelny puts it, "human cognitive competence is a collective achievement and a collective legacy; at any one moment of time, we depend on each other, and over time, we stand on the shoulders not of a few giants but a myriad of ordinary agents who have made and passed on intact the informational resources on which human lives

[118] It is an empirical question the extent to which this inheritance is the product of unconscious imitation (Henrich & Boyd 2002) vs. socially recursive inference (Scott-Philips 2015). See Sterelny (2017) for an overview of the debate.

depend." (Sterelny 2012: xii). We inherit what Quine called the "lore of our fathers" (1951): essentially long shelf-life information.[119] Each generation inherits the information and misinformation of the previous one and is able to expand upon and occasionally overhaul a subset of it because they do not need to re-discover the insights of their forebears.

Some amount of cumulative cultural evolution is possible for those with joint intentionality, since socially recursive inference is sufficient for figuring out what conspecifics are trying to do and how they are trying to do it. However, cumulative cultural evolution is significantly advanced when the game of giving and asking for reasons is introduced. When our cultural inheritance is partitioned into discrete, debatable statements, we are in a better position to determine which parts of it require revision. This, of course, requires higher-order reflection.

As I said above, inquiry requires that we provisionally assume that our foundational beliefs are for the most part true. The assumption is methodologically necessary but not dogmatic because inquiry conducted within its scope could potentially provide countervailing evidence. Furthermore, the more mature such inquiry gets, the narrower the gap between lack of positive evidential support for the assumption, on the one hand and defeat on the other. Since you inherit not only beliefs but also a practice of assigning certain beliefs foundational status from previous generations, it follows that you outsource some of the necessary inquiry to them. That is, the previous generation (often) regards the same belief types as foundational as yours does. Part of your intellectual inheritance is the progress they've made on repaying the debt incurred by the background assumptions you both share. You inherit their lore and part of that

[119] See Laland & Hoppitt (2003); Danchin & Luc-Alain (2004) and Sterelny (2012: 29).

lore is the progress they've made toward explaining how vision, for instance, is generally reliable.

Our relation to the previous generation and their lore is not entirely deferential. At a certain point our relation to them is a diachronic version of the game we play synchronically with one another. Sometimes the attempt to expand upon the discoveries of previous generations results in the acquisition of recalcitrant evidence: that is, the accumulation of defeaters. In this case we adjudicate the conflict with them just as we would one with a contemporary, except for the fact that they aren't around to advocate for themselves. So, just as in the case of private thought, we play both the part of the participant and the critic. However, instead of assessing our own thought from the perspective of the generalized other, we instead simulate the previous generation's attempt to do just this as we play the critical other. In either case, we go through the procedure laid down by defeasible reasoning: we check for *prima facie* reasons, then for defeaters, then check to see if there are defeater-defeaters and, if so, whether they are restoring or not. Because the game of giving and asking for reasons can adjudicate diachronic and synchronic conflict in the same way, it enables us to incrementally accrue cognitive capital across generations the same way we accrue it within a generation.

I submit that this is the best explanation of cumulative cultural evolution in humans. I am essentially just fleshing out the kind of explanation typically given by evolutionary psychologists/anthropologists by appealing to the defeasible reasoning tradition. This has the added benefit of giving us a more complete picture of (1).

It is now time to explain and defend,

(2a) The game of giving and asking for reasons is a coordination-conducive procedure: we lay down our cards and defer to the winning hand.

If someone offers an argument that p, there is rational pressure on their audience to either accept the conclusion or find fault with the argument.[120] You could find fault with the argument in a few ways. You could deny that it even provides *prima facie* reason to believe p. You could also try to show that even though it provides a *prima facie* reason to believe p, it doesn't provide an undefeated reason to believe p. To do this you would have to find a defeater for the reason offered. If you were unable to find one, then you would face rational pressure to accept p. If you found one and the speaker were unable to find a restoring defeater-defeater, there would be rational pressure on her to retract her assertion that p and revise her belief that p.[121] In any case, there is rational pressure on one party to yield to the other. GoGAR is conducive to belief coordination because of this. Each party is responsive to rational pressure because rational pressure is a standing within GoGAR. Furthermore, their own individual thought is regulated by internalizing the standards of GoGAR and self-monitoring for discrepancies between the standards and their actual beliefs. This is what higher-order epistemic reflection involves.

Finally, that gets us to,

(3) For those enjoying collective intentionality, private individual thought is the interiorized form of the game of giving and asking for reasons.

[120] For more on rational pressure, see Goldberg (2020).
[121] Cf. MacFarlane (2014: 109).

The game is interiorized when a single person plays both gives reasons and asks for them. In the interpersonal form of the game, at any given point one person will be giving reasons why we should believe what she recommends, and the other person will be playing a critical role by asking for reasons and assessing the quality of the reasons she gives. In the interiorized form of the game, a single person plays both roles. Tomasello illustrates the point as follows,

> Thinking would seem to be a solitary activity. And so it is for other animal species. But for humans, thinking is like a jazz musician improvising a novel riff in the privacy of his own room. It is a solitary activity all right, but on an instrument made by others for that general purpose, after years of playing with and learning from other practitioners, in a musical genre with a rich history of legendary riffs, for an imagined audience of jazz afficionados. (Tomasello 2014: 1).

In the solitary case, the subject only has her own total evidence to work with, both in making arguments and in assessing them. This is what explains the significance of justification as understood by the defeasible reasoning tradition. According to the defeasible reasoning tradition, justification is the standing one has when one's *prima facie* reasons are undefeated by one's total evidence. Individual thought is itself the solitaire version of the game of giving and asking for reasons, so justification is determined by the subject's standing in a version of the game where the deck is restricted to her internally accessible evidence.

4.4 Conclusion

We now have a naturalistic account of epistemic reflection. To achieve this, we needed to first consider what differentiates thought from mere information processing. The difference is cognitive self-regulation. We then considered qualitatively different kinds of cognitive self-

regulation that emerge when the most basic regulatory capacities take a social turn by interiorizing increasingly complicated forms of social cooperation. We found that our own form of cognitive self-regulation is what emerges from the interiorization of the game of giving and asking for reasons. In doing so, we managed to synthesize the first-personal approach to epistemology common among internalists and the third-personal approach preferred by externalists. In our account of the normative kinematics of GoGAR, we considered the defeater-checking procedure a participant would employ as she makes moves within the game. This involves considering the game from the first-person perspective of a participant. However, we also considered the truth-conduciveness and self-correcting potential of GoGAR itself from a third-person perspective not unlike the perspective an engineer might take toward her invention. In doing so, we secure the truth-connection that people often worry goes missing when we take the first-person perspective.

Chapter 5: Cognitive Autonomy

Virtue epistemologists often claim that an individual's belief must be suitably integrated into her wider cognitive architecture for that belief to count as knowledge.[122] After motivating this claim, I will argue that we can't fully understand cognitive integration in the individual without also understanding the social version of cognitive integration. Just as the individual must integrate the outputs of various information-extraction processes into a single, agent-level doxastic response, social groups must adjudicate conflicts between group members' beliefs and integrate them into a single, group-level doxastic response.

Work in cognitive science (discussed in the previous chapter) claims that individual thought in cognitively mature humans is what happens when the exchange of reasons necessary for interpersonal cognitive integration is interiorized by the subject.[123] Call this the "social intentionality hypothesis". According to the social intentionality hypothesis, we achieve cognitive integration at the interpersonal level by playing the game of giving and asking for reasons. We achieve cognitive integration at the intrapersonal level by playing the solitaire version of the game of giving and asking for reasons. The upshot is that we can kill two birds (i.e., solve two integration problems) with one stone (i.e., the capacity to make and assess arguments).

The picture that emerges brings together insights from virtue epistemology and reasons-first epistemology. Cognitive integration is necessary for knowledge. Cognitive integration

[122] See Greco (2010), Pritchard (2012) and Carter (2022a, b).
[123] See for example Billig (1996), Frawley (1997), Dessalles (2007), Hrdy (2009), Mercier & Sperber (2011; 2017), Godfrey-Smith & Yegnashankaran (2011) and Tomasello (1999; 2014; 2021). Aspects of Karmillof-Smith's (1992) view are in the neighborhood.

requires an agent-level capacity for executive functioning. For cognitively mature humans, this takes the form of interiorizing the social practice of giving and asking for reasons.

I will argue that the above enables us to better understand a range of epistemic phenomena. First, it helps us understand the true significance of Sosa's (1991; 2007) distinction between animal knowledge and reflective knowledge. Relatedly, it puts us in a position to appreciate the insights and blind spots of "small children and animals" objections to cognitively demanding epistemic requirements on knowledge.[124] Very briefly, knowledge of any kind requires cognitive integration. However, integration takes different forms in different creatures. In humans, it takes a different form in early ontogeny than it does in adulthood.[125] Lastly, it enables us to develop a virtue epistemology that gets around the passive knowledge problem raised by Baehr (2006) and Lackey (2009).

In section (I) I draw on work in virtue epistemology to motivate the need for cognitive integration in the theory of knowledge. In section (II) I draw on work in cognitive science to motivate a taxonomy of thought. Full-fledged thought requires a capacity for executive functioning. I will consider three forms of executive functioning and explain how each form builds upon the capacities at play in the previous one by interiorizing a more complicated form of social cooperation. The final form involves regulating one's own cognition by interiorizing through offline simulation the critical perspective of an imagined interlocutor. In section (III) I discuss how the findings of the previous sections favor reasons-first virtue epistemology. I then apply reasons-first virtue epistemology to the topics mentioned above.

---

[124] See for example Audi (1993), Lackey (2005), Moon (2018).
[125] See Tomasello (1999; 2003; 2014; 2021).

5.1 Cognitive Integration and Virtue Epistemology

I will begin with a case Keith Lehrer used to illustrate the problems that beset simple versions of process reliabilism.

> TRUETEMP: Suppose a person, whom we shall name Mr. Truetemp, undergoes brain surgery by an experimental surgeon who invents a small device which is both a very accurate thermometer and a computational device capable of generating thoughts. The device, call it a tempucomp, is implanted in Truetemp's head so that the very tip of the device, no larger than the head of a pin, sits unnoticed on his scalp and acts as a sensor to transmit information about the temperature to the computational system in his brain. This device, in turn, sends a message to his brain causing him to think of the temperature recorded by the external sensor. Assume that the tempucomp is very reliable, and so his thoughts are correct temperature thoughts. All told, this is a reliable belief-forming process. Now imagine, finally, that he has no idea that the tempucomp has been inserted in his brain, is only slightly puzzled about why he thinks so obsessively about the temperature, but never checks a thermometer to determine whether these thoughts about the temperature are correct. He accepts them unreflectively, another effect of the tempucomp. Thus, he thinks and accepts that the temperature is 104 degrees. It is. Does he know that it is? (Lehrer 1990: 162-3)

The intuitive answer here is "no". At first, it might seem that there is a straightforward evidentialist vindication of the intuitive verdict: the subject has no evidence that these beliefs are formed by a reliable process, so they are not justified (either propositionally or doxastically) and consequently their belief is not knowledge. As Carter (2022a: 23) points out, this response might

work for Lehrer's original version of the case. However, we can change the details slightly and the problem rears its head in a different form. Suppose the subject has a tempucomp$_{\text{DELUXE:}}$ a device that compels the subject to form true beliefs about the temperature and about the reliability of the device. We can stipulate that these beliefs are justified. Suppose further that the device hypnotizes the subject into basing her belief about ambient temperature on the belief about the reliability of the device. In this version of the thought experiment, the subject has good evidence, and her belief is based on that evidence. The problem is that she was just a bystander. The device gathered the evidence, implanted it, and initiated the basing process. If her own cognitive faculties had overseen the process, the result would have been knowledge. However, the device oversaw the process, replacing her cognitive faculties.

Here we have reason to look to (faculty) virtue-theoretic epistemology for help. According to this order of explanation, knowledge is understood as cognitive success attributable to cognitive ability. Two qualifications are in order. The first is that there are two intuitions about knowledge that guide virtue-theoretic approaches, and it is important to distinguish them. Following Pritchard (2012) we can call the first one the "anti-luck intuition". This is the intuition that at least some kinds of epistemic luck undermine knowledge. This intuition is what drives the Gettier problem. It is an intuition shared by virtue epistemologists and most non-virtue epistemologists alike. The second intuition, the "ability intuition", is the intuition that knowledge entails cognitive ability on the part of the knower. Some virtue epistemologists think that the same conditions on knowledge can vindicate both the anti-luck intuition and the ability intuition (e.g., Sosa 2007, Greco 2003; 2010). Pritchard (2012) is skeptical and sees them as two independent constraints requiring separate treatment. My proposal here involves going in for

virtue epistemology to satisfy the ability intuition. I argued in the first chapter that it is ill-suited to vindicate the anti-luck intuition.

The ability intuition has it that knowledge is true belief attributable in the right way to cognitive ability. The virtue-theoretic order of explanation involves explaining knowledge in terms of cognitive ability. Some virtue-theorists think that the cognitive abilities themselves can't be characterized without reference to knowledge. Call these people "knowledge-first virtue epistemologists".[126] Both Pritchard and I, albeit in different ways, hope to show that we can give a reductive or analytical account of knowledge. This requires that we characterize the abilities themselves without reference to knowledge. As Pritchard puts it,

> Given that we are engaged in the analytical project it is clearly not ultimately helpful to characterize cognitive abilities in terms of their conduciveness to knowledge. We thus need to ask what it is about the particular belief-forming dispositions that qualify as cognitive abilities that makes them knowledge conducive. I take it that as a minimal requirement these belief-forming dispositions should be both reliable and suitably integrated with the agent's other belief-forming dispositions. The former requirement is needed if we are to think of these dispositions as genuinely akin to skills or abilities more generally, while the latter requirement is needed if we are to think of these dispositions as genuinely reflecting the subject's cognitive agency. (Pritchard 2012: 262).

In the above quote Pritchard gives us two criteria for cognitive abilities. The first is that they are reliable. The second is that they are suitably integrated with the agent's other belief-

[126] See Mirrachi (2015) and Kelp (2016; 2018). See Simion (2018) for a similar view.

forming dispositions. The first criterion is necessary (but not sufficient) because the virtue-theorist claims that knowledge is true belief attributable or creditable to the subject. Reliability provides the truth-connection necessary for the fact that the subject believes truly to be attributable or creditable to her. However, the ability doesn't just need to be suitably connected to the truth: Tempucomp is reliable. For it to be an ability of the subject, it must also be integrated into her wider cognitive architecture. On Pritchard's view, the cognitive integration requirement ensures that the reliable process reflects the agency of the subject.

The problem with tempucomp$_{\text{DELUXE}}$ is that the true beliefs it produces don't reflect the agency of the subject. The device is not integrated with her cognitive faculties: it replaces them. John Greco has proposed that a necessary condition for a belief-forming process to be integrated into the subject's cognitive architecture is that the beliefs it produces are dispositionally sensitive to her other belief-forming processes (2010: 152ff). Perhaps this gets us around the problem.

Being integrated is a state. That state isn't sufficient for knowledge unless it is brought about by the subject. Otherwise, the resulting state doesn't reflect the performance of the subject. Yet in attributing knowledge we positively evaluate the performance of the subject. Adam Carter (2022a: 23ff) has made this point with further modifications of the Truetemp thought experiment. We can conceive of variations of Truetemp where the subject's beliefs satisfy Greco's condition, and a state of integration is achieved, but only because of compulsion from the device. Expanding on this line of thought, suppose that the device rewires the subject's other belief-forming processes. The final step in the re-wiring is that the device sets itself up to query

all her other modules for countervailing information before installing a new belief.[127] The device is sensitive to her other belief-forming processes in this case. Her native cognitive faculties and the device have been integrated. If the other faculties extract information relevant to the output of the device, it "listens". However, the resulting state of integration is not one attributable to the agent. Her faculties were benevolently pirated to bring that state about.

When the state of cognitive integration is brought about by the device, the agent is left out of the process.[128] So, true beliefs merely occur within her. Her cognitive faculties, to the extent they were involved in producing these beliefs, were manipulated by something alien. In attributing knowledge, we positively assess the cognitive performance of the subject. Such an assessment seems inappropriate in this case. The subject didn't perform. Rather, she was the theatre in which the device performed.

The insight that emerges is that only true beliefs creditable to the subject are candidates for knowledge and only beliefs integrated into the subject's wider cognitive architecture are creditable to the subject. Furthermore, the integration must itself be creditable to the subject. This means that the subject must have an agent-level capacity for cognitive integration to have knowledge. To understand knowledge, we must understand the capacity itself and the way a state of cognitive integration must be attributable to it. My focus in this paper will be on the first of these things.

---

[127] Here I use "module" in the biologist's sense, which does not entail that they are encapsulated or that they have "shallow outputs" as Fodor's (1983) use of the term entails. For more on my use of the term, see Barkow et al (1992), Carruthers (2006: Chapter 1), Mercier & Sperber (2017: Chapter 4).
[128] Cf. Reed (2016).

I want to understand the capacity by situating it within the philosophy of mind more broadly. In the next section, I will shed light on the capacity by drawing on work in cognitive science. The detour will be longer than I would typically allow. However, I will show in the final section that the pay-off is worth it.

5.2 Three Kinds of Executive Functioning

In the cognitive sciences, the agent-level capacity for cognitive integration is called "executive functioning" and it is often thought that executive functioning is what differentiates full-blooded thought from mere stimulus-response learning or the "innate releasing mechanisms" of early ethology.[129] Executive functioning makes it possible to solve new problems creatively by integrating the products of different information extraction processes (cf. Camp 2009). Although cognitive scientists typically don't use the term "integration" when describing executive functioning, many tasks attributed to executive functioning are relevant to the integration problem. For instance, empirical literature on executive functioning often describes cognitive inhibition and cognitive flexibility as core executive functions.[130] The ability to adaptively inhibit stimulus-driven processes to flexibly respond to the demands of the situation at hand are a key aspect of what cognitive integration requires. We have many stimulus-driven information extraction processes going on within us at any given time. We integrate them by selectively inhibiting some and ceding control of the agent-level doxastic response to others.

Executive functioning will always involve self-regulation through cognitive monitoring (Tomasello 2014: 14ff.). However, this takes different forms in different species. Since executive

[129] Something along these lines is claimed by Hurley (2003), Camp (2009) and Tomasello (2014: Chapter 1). Often it is simply assumed.
[130] Davidson et al. (2006), Garon et al. (2008), Diamond (2013).

functioning is what differentiates full-blooded thought from mere stimulus-response learning, different forms of executive functioning correspond to different forms of thought. Here I will consider three forms that executive functioning can take. Each form builds on the capacities at play in the previous form by situating them within increasingly complicated forms of social cooperation. The distinction between the three is philosophical, although we will see that it earns its keep in empirical work as well.

Recall from the last chapter that Individual intentionality differs from mere stimulus-response learning in two respects: it involves a capacity for causal inference and a capacity for cognitive self-monitoring. I will consider them in turn.

Full-blooded thought requires an ability to go beyond what is immediately perceptually available (Tomasello & Call 1997: 11; cf. Camp 2009: §3). This involves a capacity for storing a causal model of the environment and using it to simulate the outcomes of possible interventions. Individual intentionality also involves a capacity for cognitive and behavioral self-monitoring (Tomasello 2014: 14ff.). The subject must be capable of registering information about their own behavior and mental states as well as adaptively responding to this information when deliberating. The ability to simulate the outcomes of possible courses of action involves the subject being sensitive to her own simulations as well as the external features of the environment she is tracking in the simulation. The subject needn't conceptualize the simulation as a simulation but must be implicitly sensitive to the fact that it is one. This involves a capacity for uncertainty monitoring: the subject must be able to determine how well equipped she is to simulate the situation accurately offline. Along with causal inference, this capacity comprises the most basic form of executive functioning, and it is what makes subjects with robust individual

intentionality capable of "taking charge" in a way that creatures with only stimulus-response connections and fixed action patterns are not. So long as there is a capacity to reconsider one's own mental states adaptively and implicitly, the creature is capable of cognitive self-monitoring.

Joint intentionality is what emerges when executive functioning takes a social turn. When a subject monitors her own behavior from the perspective of a potential collaborator and regulates it accordingly, she engages in "second-personal self-monitoring".

This requires a capacity for socially recursive inference. For us to form a joint intention to pursue a common goal, we must each understand that the other understands that we understand that… this is what we are doing. The recursive reasoning will only ever be taken a few layers deep in occurrent mental episodes, but the creature implicitly understands that the pattern of inference can go on in the same way indefinitely (Tomasello 2014: 38; cf. Clark 1996). Socially recursive reasoning makes "common ground" between the parties to the joint activity possible. The common ground will be formed by ad hoc gestural communication and pantomiming, since discursive reasoning and communication only emerges alongside the next kind of intentionality to be considered.

Common ground makes possible another key feature of joint intentionality: second-personal self-monitoring. Since we each must know a certain amount about each other's mental states to form a joint intention in the first place, when we have a joint intention, we will also be able to understand what we are doing from the perspective of our counterpart. If you and I are foraging together, I know enough about your mental states to assess my own performance from your perspective. I can then modify my behavior in light of that assessment. Joint intentionality

does not include the ability to represent the performance discursively and explicitly. Creatures with joint intentionality can only understand this implicitly via simulated assessment from the perspective of their collaborator.

Second-personal self-monitoring introduces a new form of executive functioning, but not one with epistemological implications. Second-personal self-monitoring is the monitoring of one's own performance in a joint activity from the perspective of a potential or actual collaborator.[131] It is by its nature action oriented. However, second-personal self-monitoring is important for epistemology indirectly. The next form of executive functioning to be considered pertains not only to actions but to cognition in general and it builds on the capacities at play in second-personal self-monitoring. However, at this stage we have already seen an essentially collaborative form of executive functioning, to be contrasted with the "Machiavellian" (cf. Humphries 1976) form at play in individual intentionality. In joint intentionality, "bullies lose the power to bully" (Tomasello 2014: 36). This is because they have now internalized self-assessments from the perspective of their would-be victim, who is now a collaborator in joint activity.

There is some evidence that homo heidelbergensis had joint intentionality.[132]Although no extant species we know of has joint intentionality in adulthood, Tomasello (2003) argues that human children early in ontogeny enjoy joint intentionality and use it to bootstrap their way to the next kind of intentionality to be considered: collective intentionality. Later in the paper I will return to this point and bring it to bear on questions about the epistemology of small children.

[131] Cf. Karmiloff-Smith (1992)
[132] See Tomasello (2014: 36).

The next form of executive functioning is what emerges when self-monitoring is done third-personally. In second-personal self-monitoring, I monitor from the simulated perspective of a particular collaborator (either actual or potential). In third-personal self-monitoring, I self-monitor from an anonymous, critical perspective. I do this by simulating the arguments one could make against my beliefs (in cognitive self-monitoring) or my behavior (in behavioral self-monitoring). These arguments are available to anyone. This makes the method of coordination through argumentation generalized (Tomasello 2014: 92).

As in the case of joint intentionality, decision-making is no longer the decree of the strongest party. Rather, it is the product of coordination. However, now the coordination is discursive. At this stage, coordination goes beyond pantomiming and involves giving and asking for reasons.[133] Instead of gesturing at the antelope tracks and hoping you will recognize what I think we should do, I can now explicitly state the intention I recommend we form together and explicitly support my recommendation with reasons you could either accept or challenge.

One might worry that this over-intellectualizes our mental lives. Although we sometimes explicitly consider reasons for and against our beliefs, it might appear that this is something that happens only occasionally. Perhaps I have over-estimated the prevalence and significance of interiorized argumentation.[134]

In what follows I will sketch an account of how the influence of interiorized argumentation can outstrip the cases where the arguments are consciously entertained. Since the 1960's there has been a fair amount of research on how children regulate their own activities

[133] Cf. Mercier & Sperber (2017: 186).
[134] Thanks to Peter van Elswyk for pressing me on this.

with speech. Luria (1961) found that 2–3-year-old children often mimic adult instructions while performing an activity, but nonetheless behave in a way that disregards those instructions. At 4-5 years-old, however, they not only mimic the instructions but also coordinate their behavior with the instructions. At around the same age, children begin to regulate the behavior of others shortly after learning the instructions themselves (Ratner & Hill 1991; Foley & Ratner 1997). This makes sense: self-regulation is, in a sense, teaching oneself (Ashley & Tomaello 1998; Tomaello 1999: 193).

Here is Tomasello drawing out the consequences,

> Children thus show relatively clear evidence of internalizing adults' regulating speech, rules and instructions as they are reaching the later stages of the early childhood period. What is internalized is, as Vygotsky emphasized, a dialogue. In the learning interaction the child comprehends the adult instruction (simulates the adult's regulating activity), but she does so in relation to her own understanding—which requires a coordinating of two perspectives. (Tomasello 1999: 193).

The simulation of instruction is easily observable in early childhood because it involves overt speech. Later in life, something similar takes place, but the overt speech is stifled and replaced with inner speech. In both cases, this is most likely to happen when the subject faces difficult tasks (Goodman 1981). Quite often, the subject reaches a point in development where the task is no longer difficult and consciously rehearsed instruction is no longer necessary. Why is this? As Tomasello & Call put it, "once an organism has 'thought through' a problem, its future encounters with the same or similar problems may show insight and foresight on the immediate

perceptual level" (1997: 11). In these cases, the instructions still guide the subject. The difference is that the guidance is now "automatized" (Ibid). The stimulus-driven categorization has itself been molded by prior exercises of top-down self-regulation.

Tomasello and I think of cognitive self-regulation as a special case of this more general phenomenon. The above is far from a conclusive defense, but it shows that my position isn't committed to over intellectualization. Furthermore, we have seen the social intentionality hypothesis is a progressive research program (in the sense of Lakatos 1978). We have already seen that it has promising applications in several domains. In the next section, I will show that epistemology is among them.

5.3 Reasons-First Epistemology & Virtue Epistemology

We saw in section (I) that agent-level capacities for cognitive integration are a necessary condition for knowledge. In section (II) we saw that executive functioning is partially constitutive of thought. Different kinds of executive functioning involve qualitatively different kinds of thought. We are now in a better position to understand the insight of virtue epistemology from section (I) by using the findings of section (II) to shed light on the agent-level capacity for cognitive integration necessary for knowledge.

We saw in the last section that for creatures with collective intentionality, executive functioning takes the form of interiorized argumentation. A paradigmatically interpersonal activity takes a solitaire form (think of Tomasello's jazz musician). That interpersonal activity is itself a kind of cognitive integration. We saw that it serves the purpose of integrating clashing perspectives by submitting them to the force of the better argument. Put another way, a group with competing factions needs to settle on a single doxastic response. This is achieved by having

the factions play the game of giving and asking for reasons and deferring to the winning hand. This is not coordination by submission to the most powerful group member. Rather, it is a distinctively rational form of coordination.

Something analogous must take place at the intrapersonal level. This is in large part why executive functioning is necessary. We have seen that there are different forms that executive functioning can take. For creatures with collective intentionality, the method of rational coordination we see at the interpersonal level is interiorized through simulated self-assessments. This yields cognitive integration at the intrapersonal level. Much in the same way that clashing perspectives need to be integrated into a single doxastic response at the group level, processes of information extraction need to be integrated into a single doxastic response at the individual level. Consider the predicament of the subject looking at the Müller-Lyer illusion after having already seen it before and having heard an explanation of how the illusion is generated. The subject's perceptual system inclines her to judge that the lines are unequal, while her memory inclines her to judge that they are equal. She needs to integrate these information extraction processes into a single, agent-level doxastic response. She does this by going over the arguments that could be made on behalf of favoring either information extraction process and simulating critical assessments of those arguments.

By simulating the responses of potential interlocutors and using these simulations to guide her own doxastic responses, she kills two birds with one stone. That is, she achieves cognitive integration at the intrapersonal level by simulating the method of achieving cognitive integration at the interpersonal level. So, she forms her beliefs in a way that is tailored to solving both the interpersonal and the intrapersonal cognitive integration problems. In both cases,

integration is achieved not simply by one party or belief-forming process dominating the others through an excess of force. Rather, integration is achieved by submitting to the force of the better reason.

The picture that emerges combines the strengths of the reasons-first and virtue-theoretic approaches to Epistemology. We can't understand knowledge without understanding cognitive integration at the intrapersonal level. We can't understand cognitive integration at the intrapersonal level without understanding cognitive integration at the interpersonal level, at least for reflective creatures like us. Cognitive integration at the interpersonal level is achieved through the game of giving and asking for reasons. Cognitive integration at the intrapersonal level involves the interiorization (through simulated self-assessment) of the game of giving and asking for reasons.

Although we can only understand the agent-level capacity for cognitive integration in terms of the game of giving and asking for reasons, we can't fully understand that game without understanding the agent-level capacity for cognitive integration. The reason is that the point of the interpersonal form of the game of giving and asking for reasons is truth-conducive generalized belief coordination, as we saw above. The game can only serve this function if the subjects playing it can regulate their beliefs in accordance with the norms of the game.

It follows that the agent-level capacity for cognitive integration of interest to virtue epistemologists can't be understood apart from the game of giving and asking for reasons and the game of giving and asking for reasons can't be understood apart from the agent-level capacity. The two are equiprimordial for the epistemology of collective intentionality. Knowledge requires

cognitive integration that is brought about by a virtue or capacity for cognitive integration. This capacity is the capacity for executive functioning. In creatures with collective intentionality, executive functioning takes the form of self-regulation through the interiorization of the game of giving and asking for reasons.

However, for other animals this is not the case. We have seen in the last chapter that chimpanzees, for example, are capable of executive functioning but they have individual intentionality rather than collective intentionality. So, their knowledge is different than ours. It still requires cognitive integration, but cognitive integration for them involves a different kind of executive functioning. It requires an ability to adaptively respond to differences in certainty. However, this does not require that the subject be capable of engaging in or simulating interpersonal argumentation.

We are now in a position to fully appreciate the significance of the distinction between animal and reflective knowledge (cf. Sosa 2007: 34). Ernest Sosa thinks of animal knowledge as apt belief. Apt belief is accurate because adroit. Accurate beliefs are true beliefs and adroit beliefs issue from virtues or competences. So, apt beliefs are accurate, and their accuracy is attributable to the subject's virtues. Reflective knowledge is apt belief aptly noted (Sosa 2007: 43, 113). In addition to the requirements for animal knowledge, the subject must meet the further requirement of forming an apt second-order belief about the aptness of the first-order belief that satisfies the requirements for animal knowledge.

Sosa is right to distinguish the knowledge of non-reflective animals from the knowledge of their reflective counterparts. We are now in a position to see how it tracks a distinction

between two kinds of thought. Full-blooded thought requires a capacity for executive functioning, different forms of executive functioning yield different forms of thought. Creatures with individual intentionality have a form of executive functioning that does not require explicit higher-order awareness of the first-order belief forming processes being monitored. So, they can achieve the kind of cognitive integration necessary for knowledge without forming any second-order beliefs. Their form of executive functioning yields their kind of cognitive integration, and their kind of cognitive integration is the only kind of cognitive integration necessary for their kind of knowledge. Animal knowledge carves animal intentionality (i.e., individual intentionality) at the joints.

However, reflective creatures have a different form of executive functioning that requires a capacity for higher-order epistemic reflection. Interpersonal argumentation requires the ability to discursively represent the contents of first-order beliefs and their inferential relations. Executive functioning for creatures with collective intentionality involves the interiorization of interpersonal argumentation. So, their kind of executive functioning, and consequently cognitive integration, entails a capacity for higher-order epistemic reflection. Since knowledge requires cognitive integration and their kind of cognitive integration requires higher-order epistemic reflection, their kind of knowledge requires higher-order epistemic reflection. Reflective knowledge carves reflective (i.e., "collective") intentionality at its joints.

So far, we have vindicated the significance of Sosa's distinction by showing how it carves two forms of thought at their respective joints. However, it is important to note two things. The first is that we haven't yet vindicated every detail of Sosa's analyses of animal and reflective knowledge. Rather, we have seen the need for a distinction between two kinds of

knowledge, one that does not require a capacity for higher-order reflection and another that does. However, this does not entail that reflective knowledge is apt belief aptly noted.[135]

Furthermore, none of the above vindicates the epistemic significance of animal knowledge for reflective creatures.[136] We have seen that animal knowledge carves individual intentionality at its joints and reflective knowledge carves collective intentionality at its joints. This is why each kind of knowledge matters for its respective subjects. However, since animal knowledge does not carve our kind of intentionality at its joints, it is unclear what its epistemic significance is for us.

It is worth noting the implications for the epistemology of small children now that animal knowledge has been covered. It is not uncommon for people to charge theories of knowledge and justification with overintellectualizing their subject matter when they require cognitive abilities not possessed by small children.[137] Children certainly seem to know things, so we should develop a theory that either explains how or explains away the intuition.

To keep this discussion brief, I will focus on Andrew Moon's (2018) version of the small children objection because he is specifically concerned with the implications of Lehrer's Truetemp case, as are we. He considers two ways of vindicating the no-knowledge verdict. The first is that the subject lacks higher-order evidence indicating the reliability of her spontaneously formed beliefs about ambient temperature. On this proposal, the subject needs higher-order evidence that a faculty is reliable to form knowledgeable beliefs using it. The second is that the

[135] For criticism of this analysis, see Kornblith (2009). See Carter & McKenna (2019) for a reply.
[136] A similar idea is pursued by Goldberg & Matheson (2020).
[137] See for example Audi (1993) and Lackey (2005).

subject has a defeater for her spontaneously formed beliefs about ambient temperature. On this proposal, a subject needn't have any higher-order evidence that a faculty is reliable to rely on it. Rather, she need only be without any reasons telling against the reliability of the faculty to form knowledgeable beliefs using it.

We needn't worry about what exactly the defeater might be in this case, I will simply assume that there is one. What matters for our purposes is that he argues that the no-defeater view outperforms the higher-order evidence view because the former, but not the latter, allows small children incapable of higher-order reflection to acquire knowledge by relying on their cognitive faculties.

Moon's argument has both an insight and a blind spot. The insight is that a view that precludes knowledge in small children is false. The blind spot is that knowledge in cognitively mature adults and knowledge in small children is qualitatively different because the former exhibit collective intentionality and the latter exhibit either individual or joint intentionality (more on this below). We have seen that different kinds of executive functioning involve different forms of cognitive integration. Knowledge requires cognitive integration and different kinds of knowledge require different kinds of cognitive integration, as we saw above. So, for all Moon has said, the higher-order evidence view might still be right for adults but wrong as an account of knowledge in general. I don't endorse the higher-order evidence view or the no-defeater view for that matter. My own view is that the matter is complicated in ways that neither view fully appreciates, though I don't have the space to explain why here.[138] What matters here

[138] See my account of the foundations of knowledge in the last chapter though for some sense of what I have in mind.

is that we should be more careful about how we use objections having to do with small children and animals. Both adults and children have knowledge, but we shouldn't assume knowledge comes to the same thing in both cases.

Now that I have explained why it matters, I will say a bit more about the claim that small children have either individual or joint intentionality and adults have collective intentionality. Tomasello (2003) argues that language acquisition depends on the capacities for socially recursive inference discussed above in the section on joint intentionality. Studies in developmental psychology confirm that as early as 9-12 months of age children can jointly attend to objects with others (Bakeman & Adamson 1984), direct the attention of others to distal objects (Bates 1979) and follow the gestures of others (Corkum & Moore 1995). These are precisely the capacities we discussed earlier when considering how a common ground can be formed non-discursively. Gesturing and pantomiming produce what Tomasello calls "joint attentional frames" (2003). When a child becomes capable of these activities, she moves from individual to joint intentionality. Furthermore, by engaging in these mind-reading activities, children can gradually acquire a language by figuring out the communicative intentions of those around them. The acquisition of language moves them from joint intentionality to collective intentionality. The research program of cognitive functional linguistics tries to spell out the details.[139]

The last thing to point out here is that the picture that emerges from this discussion helps us account for the insights of virtue epistemology having to do with cognitive integration while

[139] See Langacker (1987; 1991; 2000), Croft (2001), Bybee (1995; 2002) and Tomasello (2003).

bypassing some of the difficulties that have been raised in the literature. For instance, Jason Baehr (2006) and Jennifer Lackey (2009) have objected to virtue epistemology on the grounds that passively formed beliefs (e.g., on the basis of testimony) can be candidates for knowledge, although the fact that the subject believes truly is not creditable to her in any interesting sense. However, if we relax the requirements for credit enough to allow for passively formed beliefs to be creditable to agent, then they will be too permissive because Gettier cases will satisfy the requirements as well (Lackey 2009).

However, I am offering a different kind of virtue epistemology where this problem doesn't arise. On the view that I recommend, the fact that the subject believes truly does not need to be attributable to her own cognitive virtues. Rather, only her state of cognitive integration needs to be attributable to her. As mentioned above, I have not yet offered an account of the relation that needs to hold between her capacity for executive functioning and the state of cognitive integration for the latter to be attributable to the former. Such an account will ultimately need to be given, but it is beyond the scope of this paper. What I hope to have done here is give an account of the faculty to which the state of cognitive integration must be attributable without giving an account of the "…is attributable to…" relation itself.

5.4 Conclusion

I have argued for reasons-first virtue epistemology here. A great deal of work is left to be done. I hope to have shown that the work is worth doing in light of the promise shown so far. To this end, I have shown that it enables us to address two integration problems at once. It helps with the intrapersonal integration problem familiar to epistemologists and with the interpersonal version of the same problem, which tends to receive more attention in cognitive science.

Furthermore, it sheds light on the significance of the distinction between animal knowledge and reflective knowledge. Relatedly, it uncovers both the insights and blind spots of the "small children and animals" objection to cognitively demanding epistemic requirements. Lastly, it enables us to circumvent the problem of passive knowledge that has long been a worry for virtue epistemologists.

Chapter 6: Epistemic Reasons: The Copernican Turn

I'll conclude the dissertation by looking into the question of what epistemic reasons (metaphysically) are. So far, I have only talked about how they are epistemically assessed and how the role that form of assessment plays in both our interpersonal and private mental lives.

When people pursue this question, they often take it for granted that it can only be answered by identifying epistemic reasons with something else. Jerry Fodor said, "If aboutness is real, it must be really something else (1987: 97). I think it is often assumed that if reasons are real, they must be really something else. The thought is that we can better understand how reasons play their distinctive role in our epistemic lives by understanding how this thing we've identified can play that role. Sometimes this kind of thinking is made explicit in the closely related debate about the metaphysics of evidence. In a recent volume on the topic, Veli Mitova frames the discussion by asking "What kind of ontological beast" (2018: 1) evidence is.

I'm going to argue that this presupposition has occluded a simple but overlooked possibility: reasons are just episodes of reasoning.[140] We shouldn't assume we are looking for an "ontological beast" we could identify under another guise (one that makes no reference to reasons or reasoning). We can nonetheless account for the considerations that motivate rival positions that share this presupposition. These include,

1) Reasons play a role in explanation.[141]

[140] Simion (2021) says that when we reason we are making a move in a practice of inquiry but doesn't take the further step of identifying reasons themselves with moves in the practice.
[141] See Williamson (2000: 195), Dougherty (2011), Littlejohn (2012a: 192).

2) Reasons (by their very nature) rationalize belief.[142]

3) Reasons (by their very nature) objectively favor belief.[143]

4) Reasons are amphibious (i.e., they have a foot in the mind and a foot in the world).[144]

Furthermore, we can do it without abandoning monism about reasons: we can hold on to the thesis that reasons are a unified class rather than an ontological medley of disparate entities.[145] Furthermore, reasons, on this account, aren't any more mysterious than the familiar activity of reasoning itself.

Most importantly, however, we can get all this by turning the arguments that have been put forth in favor of rival views against them. Here is why: most of the arguments put forth in their defense don't tell us anything about what reasons are. Rather, they tell us about how reasons are either specified or epistemically assessed. What is being specified or epistemically assessed will in every case be an episode of reasoning. So, that is what reasons metaphysically are. This should not be surprising. As Schroeder has pointed out, reasons are competitors (cf. Schroeder 2021: Chapter 2).[146] Reasons are, by their very nature, the kind of thing that can be rebutted or undercut; they are essentially subject to challenge. It follows that to understand reasons *as* reasons is to situate them in competition with one another. The activity of reasoning is

---

[142] See for example BonJour (1985), Fumerton (1995), Conee & Feldman (2004), Goldman (2009), Swinburne (2011), Neta (2018), Schroeder (2021: 66). Turri (2009) and Gibbons (2010) say that reasons make beliefs reasonable, which may or may not come to the same thing.
[143] Littlejohn (2012a: 147, 228), Mitova (2015). Parfit (1997), Dancy (2002), and Broome (2004) make a similar point with respect to reasons for action.
[144] See for example McDowell (1994), Neta & Pritchard (2007), Neta (2008; 2018), Pritchard (2012; 2018), Gibbons (2013), Mitova (2015), Simion (2023).
[145] For a pluralist view that denies this, see Kelly (2008), Rysiew (2011; 2018) and Fogal (2018). I will discuss the position more below.
[146] A similar point has long been recognized regarding practical reasons. See for example Ross (1930), Broome (2004), Lord & Maguire (2016), and Snedegar (2021).

that competition. It is the activity in which reasons are put forth and jeopardized by other reasons. I will argue that rival accounts provide many insights that need to be preserved, but they don't give an adequate account of the competitive nature of reasons.

The upshot is a Copernican turn in the metaphysics of reasons. Instead of trying to better understand epistemic reasons by identifying them with something else in our ontology, we understand the epistemic significance of other things in our ontology in terms of how those things bear on the epistemic assessment of reasoning.

To make my point about reasons, I will often consider arguments people have made about evidence. If not identical, reasons and evidence are very similar phenomena and similar arguments can be made about them. This is reflected by the fact that some philosophers use the terms interchangeably.[147]My own view is that "evidence" has several senses, one of which is synonymous with "reason". Partly for this reason, I will use the terms interchangeably as well. Going forward, I will use "reason" interchangeably with "reason to believe". I think something like what I say about reasons to believe probably applies to reasons for action.[148] However, I will restrict my focus to reasons to believe here. Otherwise, the amount of literature I would need to engage with would be unmanageable for a single paper.

Here is how I will proceed. In section (I) I will give a brief sketch of the kinematics of the activity of reasoning. Here I will pre-empt some concerns about the project that I suspect have

[147] For example, Conee & Feldman (2004: 19; 296), Dougherty (2011). Schroeder (2021) makes a compelling argument that we should identify evidence with reasons.
[148] Hieronymi (2021) has argued for a reasoning first account of practical reasons, but she takes the relation to a question as basic whereas I take the norms of reasoning as basic. Accordingly, she offers a much different argument than the one pursued here.

prevented others from pursuing it. I won't provide the entire positive argument for the view here. Much of the argument will come in the following sections. Here I will put the pieces in place to make that argument later. In section (II) I will consider Williamson's (2000) argument that evidence must be propositional because otherwise there would be no explanation of (1). Conee and Feldman (2011) nicely diagnose a legitimate concern about his argument: while it is true that in explaining something we cite propositions, it doesn't follow that either the thing explained or the thing doing the explaining is itself a proposition. In the remaining sections I argue that other views (including that of Conee and Feldman) are open to a criticism along the same lines.

In section (III) I will make a similar point against the view that reasons consist in (non-factive) mental states or their contents. Proponents of this view claim that non-factive mental states are uniquely well-suited to make sense of (2). In a way, this is right, but it only tells us about how reasons are assessed and not what they are. In section (IV) I will consider the analogous problem for proponents of factive reasons. Proponents of factive reasons think that only facts can make sense of (3). In a way this is true, but it has more to do with how reasons are assessed than what they metaphysically are. In section (V) I consider views according to which reasons are amphibious, in the sense that they have a foothold in the mind and the world. Here I will argue that the epistemic assessment of reasons is amphibious: reasons are assessed both in light of the facts and the subject's mental states. Nonetheless, reasons themselves aren't amphibious, or at least not in the way they are often taken to be in the literature.

6.1 A Preliminary Sketch of the Account

I will begin by giving a preliminary sketch of the view. I will provide some of the arguments for it in this section, but most will appear in later sections. The impulse driving this

paper is ecumenical. Although I disagree with the details of all the views to be considered, I am sympathetic to the spirit of each. Indeed, much of the appeal of the view I recommend is that it mimics the things other views do well while avoiding familiar objections. To demonstrate this, I will need to say a bit about what the view is first, dispel some likely misunderstandings, and then compare it to the views already out there. Here it goes.

Reasoning is like other activities with constitutive rules. The constitutive rules determine both what counts as reasoning (at all) and what counts as good reasoning. I will treat these in order.

Call the kind of belief that is formed for reasons a "reflective belief". When I say that this kind of belief is formed for a reason, I don't mean that every token of it is. Part of what makes particular reflective beliefs tokens of the reflective belief type is that they are the kind of state where it makes sense to ask the subject why she is in that state. I think reflective beliefs also have a causal role that is part of their nature, but I won't consider it here except insofar as it is relevant to the basing relation. The important thing is that when a subject with a capacity for critical reflection forms a belief, it is appropriate to ask what her reason for that belief is. The appropriateness of that question is essential to the belief states of creatures with a capacity for critical reflection.

At any rate, in asking why someone believes something, you are asking for a normative epistemic reason. A mere causal explanation (i.e., an explanatory reason that is not also a normative reason) doesn't answer the question, which I will call "Anscombe's Question" going forward. Much as Anscombe (1957) claimed that intentional action can't be understood apart

from a certain kind of "Why?"" question, I claim that reflective beliefs can't be understood apart from a certain kind of "Why?" question.

Reflective beliefs are states that invite Anscombe's Question. This is, on my account, because they are episodes of reasoning. Reasoning is an activity. My claim then is that the rules of the activity are such that the formation of a reflective belief counts as an episode of it. Compare: moving your pawn forward two spaces at the beginning of the game counts as an opening in chess. The rules of chess are such that they determine what counts as playing chess: *mutatis mutandis*, on my account, for the rules of reasoning.

On the view I recommend, however, it is not the case that we first understand what reflective beliefs are and then explain episodes of reasoning in terms of this prior understanding. Rather, reflective beliefs can't be understood apart from Anscombe's Question and Anscombe's Question only arises for a state because it is appropriate to evaluate that state by the standards appropriate for assessing episodes of reasoning. You might be able to specify the neural realization of a reflective belief without reference to that activity, much as you could specify the material constitution of what is in fact a pawn without saying anything about chess. But doing so would tell you nothing about what pawns (as such) are. The same goes for reflective beliefs and their neural realization. Consequently, neither reflective beliefs nor the standards of the activity admit of a reductive analysis. Rather, we must make sense of both at the same time by situating them in a common network.[149]

[149] That is, I am offering what Kelp (2021) has called a "network analysis", rather than a "dismantling analysis".

The upshot is that reflective beliefs can only be understood as the kind of thing they are in terms of the rules of the activity. Put another way, you only understand reflective beliefs by understanding them as the kind of state subject to a certain form of epistemic assessment. The constitutive rules of the activity specify that the formation of a reflective belief is what counts as an episode of the activity. A reflective belief is the kind of thing subject to the standards of epistemic assessment laid down by the rules. This is not circular. Two different aspects of the rules are in play. The first clause of the rules specifies that reflective beliefs are the things that count as episodes of reasoning. Reflective beliefs are defined in terms of a further aspect of the rules: the standards of epistemic assessment that determine how episodes are assessed.

The natural next question, then, is why certain episodes in the natural world are subject to these standards while others aren't. That is a meta-normative question. It asks about the source of the normativity of the rules. This is a legitimate question, and it must be answered sooner or later. However, I won't try to answer it here. The purpose of this paper is to explain the ontological status of reasons in terms of the activity of reasoning and the standards of epistemic assessment proprietary to it. The question of how those standards become normative for us is another question. What I say in this paper is compatible with several answers to it. These norms could be grounded in something like an epistemic social contract theory or by social conventions of some kind.[150]They could be grounded in epistemic reactive attitudes.[151]Perhaps it is simply a

[150] Dogramaci (2013) gives an account along these lines for the norms of deductive inference. Sellars (1956/1997) and Searle (1995) give different accounts that fall into this category.
[151] See for example Brandom (1994).

brute fact that the activity has certain constitutive norms.[152] I also leave open the possibility that a proper-functionalist explanation will prove to be correct.[153]

I am not the first person to claim that epistemic norms are (at least partially) constitutive of belief.[154] Sometimes people object to other ways of pursuing this idea on the grounds that it is not naturalistic.[155]Whether this is true depends on what counts as "naturalism", which I won't try to settle here. Regardless of whether other people who have said this are naturalists, the view I adumbrate here is consistent with naturalism on any plausible understanding of the term. Although I say that epistemic norms are partially constitutive of reflective belief, I have mentioned various meta-normative accounts compatible with the claims in this paper. Most of the options I mentioned explain the source of the norms in clearly naturalistic terms.

Now it is time to say a bit about how the episodes of the activity are assessed. I think the kind of assessment at issue has two dimensions: the subject's reasons are assessed both relative to the facts and relative to her total evidence. When they are assessed favorably along the former dimension, they are objective normative reasons. When they are assessed favorably along the latter dimension, they are subjective normative reasons.[156] Reasons can be subjective normative without being objective normative. This happens, for example, in envatted brain cases. Reasons can also be objective normative without being subjective normative. This happens when you have a *prima facie* reason to believe something, but misleading evidence that defeats it. There is

---

[152] Cf. McDowell (1994).
[153] Millikan (1984), Burge (2010), Graham (2012) and Simion (2018) all think epistemic normativity is grounded in proper cognitive functioning.
[154] See Sellars (1956/1997), Davidson (1980), Dummett (1983), Brandom (1994) and McDowell (1994), for example.
[155] For example, Carruthers (2006).
[156] So, the constitutive rules are of the non-imperatival variety identified by Searle (1969: 34ff.).

a fact out there that defeats the defeater, but it is beyond your ken. In cases like this, your reason is objective normative, but it nonetheless doesn't justify your belief. It might be a good reason, but it is not a good reason *for you*. This is not uncontroversial. However, it is (for many) an intuitive starting point. After fleshing out the view a bit more, I will use the picture that emerges to address some of the arguments that animate the controversy.

I will not attempt to give necessary and sufficient conditions for favorable assessment along either the subjective or objective dimension. I will, however, assume a few things that a complete account will have to either explain or explain away. The first is that reasons are, by their very nature, competitors.[157]That is, they are in competition with one another to tip the rational balance (i.e., total weight of reasons) one way or another. One reason can outcompete another by being a weightier reason against belief than the first one was for belief. One reason can also outcompete another by undercutting or attenuating the second reason. These phenomena have been extensively explored in the defeasible reasoning literature. I take it the leading proposals to emerge from it are at least roughly extensionally adequate when it comes to accounting for when a subject's reason is sufficient relative to a body of other, potentially countervailing, reasons.[158] When the subject's reason is sufficient relative to her total evidence, she has a subjective normative reason and when her reason is sufficient relative to the facts, she has an objective normative reason.

[157] Cf. Schroeder (2021: Chapter 2).
[158] See Lehrer (1965; 1970), Lehrer & Paxson (1968), Hilpinen (1971), Johnsen (1974), Swain (1974), Barker (1976), Klein (1971; 1976; 1980), Pollock (1986), Moser (1989), Audi (1993), Schroeder (2015). See de Almeida & Fett (2016) for a response to criticisms of defeasible reasoning. See Shope (1983) for a useful overview.

Before going on, it is important to note that the Copernican Turn I recommend is not a form of idealism. The facts aren't noumena beyond the activity. It is not the case that good reasons (in either the subjective normative or objective normative sense) are just whatever our epistemic community will let us get away with saying.[159] Rather, as we have already seen, the epistemic assessment of the episodes is determined (along the objective normative dimension) by the facts. Compare the rules of baseball. If the trajectory of a pitch crosses home plate within the strike zone (i.e., above the batter's knees and below the letters of the jersey), then it is a strike. A fully objective fact determines whether the pitch is a strike. The umpire can't make a pitch a strike; that is why it makes sense to sometimes disagree with the umpire's call. The umpire can make it the case that a pitch is a strike *for the purposes of MLB scoring*, however. Relatedly, just as whether a pitch is a strike is determined by a fully objective fact about its spatial trajectory, whether a subject's reason is objective normative is determined by fully objective facts and not by what any particular scorekeeper takes those facts to be. Similarly, whether her reason is subjective normative is determined by facts about her total evidence and its rational relations to the belief for which it is a reason, independent of what any scorekeeper happens to think.[160] The correct assessment of the episodes of the activity is like the correct assessment of balls and strikes and not like the assessment of balls and strikes by decree of the umpire for the purposes of MLB scoring.

I will conclude this section by tying the above bits together to give a fuller picture of how the activity works. As I said, forming a reflective belief is an episode of the activity. It is possible

[159] As Rorty (1979) famously claimed.
[160] Cf. Swinburne (2011). For a dissenting view, see Foley (1993).

to form a reflective belief for no reason at all, although such beliefs are assessed negatively along both the subjective and objective dimensions. Doing well in the activity requires that you form your belief for sufficient reason (though not necessarily in an act of conscious awareness). To form your belief for a sufficient reason, you must form it for a reason.

Suppose you base your reflective belief that the cat is orange on your perceptual experience as of an orange cat. That is an episode of the activity. When you have a perceptual state, you have a *prima facie* reason to believe things suitably related to its content. The same goes for memory and perhaps testimony as well. I say that the episode itself rather than the perceptual state is your reason. Reasons are competitors by nature and your perceptual state is not a competitor.

I anticipate resistance on this point, so it is worth slowing down. One might insist that perceptual states tip the rational balance, so they are competitors. They are in a competition because other things can tip the rational balance the other way. People often liken reasons to marbles on a scale and say that they are in competition in much the same way the marbles are (cf. Broome 2004; Snedegar 2021). The metaphor helps illustrate its own shortcoming. Are the marbles themselves competitors? Is competition part of their nature? Clearly not. The marbles don't have any tendency to oppose each other except insofar as we put them on a scale. Their weight determines how they are disposed to affect it. However, the mere fact that some marbles weigh more than others doesn't mean that there is any competition between them. It is not as if there was a competition taking place before we placed them on the scale and by putting them there, we learned which marbles (already) won.

The same applies to perceptual states. Perceptual states might tip the rational balance, but this doesn't make them any more competitive by nature than the marbles. For the marbles to oppose each other in any sense, we have to put them on a scale. For perceptual states to compete with other considerations, we have to do something with them: we have to reason on their basis. It is only relative to the activity of reasoning that there is any opposition between perceptual states and other considerations. Reasoning is, by its very nature, a competitive activity. This is not to say that it is by nature something done for sport. Rather, it is by nature the sort of activity such that episodes of it can either support or challenge one another. Episodes of reasoning are competitors by nature because they are moments of an activity that is competitive. Perceptual states are reasons in a sense, but it is a derivative sense. They are reasons only insofar as they figure into episodes of an activity that is competitive by nature. I will argue in the penultimate section that the same applies to facts, although facts and mental states bear on the form of normative assessment proprietary to the reasoning in different ways.

If you formed your belief by basing it on a perceptual state and I asked you Anscombe's Question, you probably would just say, "My reason is that I saw the orange cat." This might seem to commit us to the identity of the perceptual state and the reason. We will see reasons to think otherwise in the next section. For now, I will just provide an error theory. Anyone asking Anscombe's Question wanted to know about your episode of reasoning. Each episode involves a transition from something to something. Whoever was asking the question already knew one relatum of the transition. That is, they already knew that the transition is from something to the reflective belief that there is an orange cat before you. So, you can answer the question concisely and informatively by specifying the part of the transition they didn't already know about.

In each episode you regulate your own thought in accordance with what you take to be the state of the scoreboard. This is how reasons can be causes. When you believe something for a reason (the kind of reason that answers Anscombe's Question), you make a transition in the activity that you think fares well by its standards. This doesn't necessarily mean that you have a higher order thought that this is the case. Rather, in basing your belief on something, you implicitly take the transition you've made to be a good one. This is why it is incoherent to believe something for a reason you think is a bad reason.

When you believe q because you made a transition in the activity of reasoning by basing q on something, you take that transition to be normative (at least implicitly). The fact that you took that transition to be normative explains why you believe p. Making that transition caused you to believe q. Another way of putting it is that the transition you've made is your motivating reason for believing q and this just means that you took the transition to be a normative reason. The fact that you took it to be a normative reason is the explanatory reason why you believe q.[161]

There are both subjective and objective normative reasons and I deliberately didn't specify which you took your reason to be. Some philosophers think that we should just be aiming for justified beliefs (i.e., sufficient subjective normative reasons).[162] Anything more than that requires worldly cooperation beyond our control. Following Littlejohn (2012a: 2) let us call this move the "Russellian Retreat". Proponents of the retreat think that when we take our reasons to be normative, we take them to be subjective normative.

[161] Cf. Schroeder (2007)
[162] For example, Kaplan (1985), Wright (1991), Pollock & Cruz (1999).

I am skeptical, although you could agree with the key claims made in this paper and be an advocate of the retreat. I argued in chapter 2 that subjective normative reasons *just are* apparent objective normative reasons to someone with your total evidence. I will later show how this thesis can help give an error-theory of some claims made by proponents of factive reasons.

That is the sketch of and preliminary argument for the view I recommend. In the following sections, I will argue for it by trying to show that arguments for opposing views, properly understood, actually support this one.

6.2 Reasons and Explanations

Let us begin with an argument from Williamson that evidence must be propositional. He says, "Evidence is the kind of thing which hypotheses explain. But the kind of thing which hypotheses explain is propositional. Therefore, evidence is propositional" (Williamson 2000: 195).[163]

The major premise is that evidence is the kind of thing which hypotheses explain. This seems plausible enough. When you get a test result, you have evidence of your condition. Hypotheses have the potential to explain that evidence. If your evidence is that you have a fever, the hypothesis that you have the flu might explain your evidence.

The minor premise is that the kind of thing evidence explains is propositional. Notice in the example just considered that the thing explained by the hypothesis is that you have a

[163] Cf. Dougherty (2011) and Littlejohn (2012a: 99-102).

temperature. That-clauses take propositional complements.[164] Since evidence explains the complement of a that-clause, it explains a proposition (a true one).

The argument seems plausible enough at first glance, but Conee & Feldman (2011) have challenged it in the following way. It may be that when you are explaining something, you cite propositions. However, it doesn't follow that either the explanation or the thing explained *is* a proposition. Both the major and the minor premise seem plausible when we consider examples of people providing explanations. When giving those explanations, one cites propositions. So, it is tempting to think that the thing being explained and the thing doing the explaining are propositions. After all, in order to indicate what you are explaining and how you are explaining it, you have to cite propositions.

Pryor (2007) makes a similar point. He acknowledges that we use that-clauses to specify which reasons are the subject's reasons. However, he points out that it doesn't follow that the proposition cited in specifying the subject's reason is identical to the subject's reason. In fact, it gives us reason to think that the proposition cited to specify the reason is not identical to the reason. Consider the following pair of sentences,[165]

1a) You met the woman John chose. She was Kristen, wasn't she?

1b) #You met the woman John chose. She was Kristen, wasn't it?

1a sounds fine. The first sentence introduces a referent and the second sentence first states that the referent is identical to Kristen and then asks whether this is all correct. That is, the copula is

[164] See Moffett (2005) for a defense.
[165] (Ibid).

equational: it states an equality or identity. 1a equates Kristen and the referent of the definite description from the previous sentence. 1b is infelicitous. It stated the identity of Kristen and the referent of the definite description. That is, the copula is equational, as before. "She" picks up the referent anaphorically. So, "it" doesn't work. It can't refer back to Kristen, "she" is already doing that. The copula is there to state the identity of the object of the choice. It isn't there to talk about the choice itself. So, "it" can't pick up the choice itself anaphorically. Now consider the following pair,[166]

2a) John's choice was Kristen, wasn't it?

2b) #John's choice was Kristen, wasn't she?

Now "it" sounds fine, and "she" sounds bad. Why? It sounds bad because we are now using "Kristen" to specify the content of John's choice. The copula, "was", is not equational. It is specificational.[167] It isn't used to state an identity but to specify the choice John made. The anaphor refers back to the choice and the question is whether the choice was correctly specified. "She" is infelicitous because the choice doesn't have a gender. If the specification of the choice were identical to the choice itself, then both would be felicitous. But they aren't. So, we shouldn't identify the way the choice is specified with the choice itself.

A similar point applies to reasons. When we use a that-clause in attributing a reason to a subject, it seems clear that we are at least specifying the content of the reason. Why should we think we are also stating an identity between the reason and the proposition that is its content? In

[166] (Ibid).
[167] For more on the specificational copula, see Williams (1983), Partee (1986), Moro (1997), Schlenker (2003), Mikkelson (2004) and Romero (2005).

the other cases we just considered, the specificational use of the copula was not an equational use. Perhaps the case of reasons is different, but I see no reason to think so. So, I will count this as defeasible evidence of their non-identity and incorporate this point into a cumulative case.

For all I say here, it may be the case that reasons have propositional contents. In fact, I am sympathetic to this claim (although you could disagree with me about that and still agree with the main thesis of the paper). However, this doesn't mean that reasons *are* those contents. It just means that reasons are individuated at least as finely as propositions. However, they could be more finely individuated than propositions in much the same way speech-acts are, on the assumption that speech-acts are individuated by propositional content and illocutionary force.[168]

Turri (2009) has pointed out that the same proposition can be used in different ways in reasoning. If so, the subject's reasons are different than the propositions involved in her reasoning. I could believe p→¬q, suspend as to whether q, and on this basis suspend as to whether p. I could also believe p→¬q, believe q, and suspend as to whether p. The first case might be rational, but the second obviously isn't. But if reasons just are the propositional contents of the reasons, then the cases involve the same reasons. But if reasons are episodes of reasoning, then they are more like speech-acts: they are moves one might make. Something like the force/content distinction could apply here as well, as it seems to in the kind of case Turri has brought to our attention. Suspending on p is a different move than committing oneself to it, even if both states are based on the same thing. Although the moves have the same propositional

[168] The assumption has been challenged by Hanks (2015).

content, they are not the same move. So, the two cases don't involve the same reasons, and this is why the chains of reasoning are not equally rational.

Mark Schroeder (2021: 77ff.) has given us further compelling reasons to doubt that whatever we cite to specify a subject's reason should be identified with that reason itself. This is because how we specify the subject's reason is sensitive to how we evaluate that reason. If I think the subject is reasoning from a true lemma, I might say, "She believes she is ill because she drank poison". Suppose I believe she is reasoning from a false lemma: she is mistaken about having been poisoned. Then I would specify her reason as follows, "She believes she is ill because she believes she drank poison". Now suppose I think she really did drink poison, but I am wrong about that. Does it follow that I am wrong about what her reason was? It seems not.[169] I may be wrong about whether she drank poison, but it doesn't follow that I am also wrong about what her reason is for thinking she is ill. It follows that we shouldn't identify the subject's reason with the thing we cite in specifying it. The subject's reason was the same whether or not she was really poisoned. However, to specify her reason we cite her belief if she wasn't really poisoned and a fact or true proposition if she was indeed poisoned. So, the thing we cite is not the same thing as the reason itself.

What is the reason itself then? To figure that out, we need to ask what it is we were specifying and evaluating a moment ago. The answer is that we were evaluating her reasoning. We were evaluating an episode of reasoning. It consisted in a move from a claim about her temperature to a claim about the flu. To make clear to us which move she is making, she would

[169] A similar point is made by Williams (1981: 102) and Dancy (1995: 13) in the practical realm.

say something about her temperature. However, what we are evaluating is the move she has made. We aren't evaluating her temperature nor are we evaluating the proposition that she used to make clear which move she is making. So, a Copernican Turn is in order. We should stop trying to make sense of reasons by identifying them with the things we cite when we reason and, instead, make sense of the epistemic significance of those things insofar as they bear on the epistemic assessment of reasoning.[170]

One might think it follows from my account and some plausible further assumptions that reasons are mental states. If we are evaluating reasoning and reasoning is something done mentally, then it might seem that reasons simply are the mental states in virtue of which the episode of reasoning is attributable to the subject. In a way, this is right. It is certainly the case that how you reasoned is determined by which mental states you had. However, it doesn't follow that your reasons are to be identified with your mental states. I will explain in the next section.

6.3 Belief and Rationalizations

I will consider an argument meant to show that reasons are non-factive mental states. Unlike the argument from Williamson in the last section, this argument is not made explicitly in print to my knowledge. However, I think something along these lines is implicit in a large swath of internalist epistemology.[171] Here is the argument. Reasons, by their very nature, rationalize belief.[172] Non-factive mental states rationalize belief. Therefore, reasons are non-factive mental

[170] My account is in some ways similar to that of McHugh & Way (2018 a, b). Like me, they think of reasoning as an activity with a constitutive aim. However, they think that the constitutive aim is to preserve the fittingness of one's attitudes. Fittingness, on their account, is explanatorily prior to the norms of the activity that preserves it. On my account, you can only understand fittingness in terms of conformity to the norms of the activity of reasoning.
[171] Though also in Goldman (2009).
[172] Cf. Schroeder (2021). Turri (2009) and Gibbons (2010) say that what is essential to reasons is that they make things reasonable.

states. A similar argument can be made by swapping non-factive mental states for their contents.[173] The objections below will address both arguments. So, I will run roughshod over the distinction between mental states and their contents in what follows.

The first premise is eminently plausible. We get our grip on reasons at least in part by understanding the role they play in our cognitive economy. The role of rationalizing belief is part of this. The second premise also sounds rather plausible, although it is more contentious. It is widely thought that two subjects alike with respect to their non-factive mental states must also have the same reasons[174] and, consequently, the same beliefs must be rational for each of them.[175] Relatedly, two subjects with the same non-factive mental states are arguably susceptible to the same defeaters,[176] which is easiest to explain if they have the same reasons. I will take this for granted for now. I will do this because it will enable me to develop a theory that will be able to account for some of the considerations that have made others think differently, as we will see in the next two sections. Furthermore, something along these lines also seems to be what people have in mind when they argue that what evidence you have must be reflectively accessible.[177] You can't determine by reflection alone whether your mental states are veridical or true. Therefore, nothing precludes some evidence being false.[178]Having false evidence, it would seem,

[173] Cf. Dougherty (2011).
[174] See Nozick (1981: 184-5), Foley (1993: 8ff.), BonJour (2003: 185ff.), Sosa (2003: 150), Silins (2005) Huemer (2006; 2007: 48) and Turri (2009).
[175] Something along these lines seems to be believed by, for example, Cohen (1984), BonJour (1985), Conee & Feldman (2001; 2004: 101; 2008; 2011), Wedgwood (2002), Brueckner (2009), McAllister (2024).
[176] Conee (2007: 19).
[177] Cf. Silins (2005), Schiffer (2009), Smithies (2012; 2014).
[178] See for example Conee & Feldman (2008), Goldman (2009), Rizzieri (2011) and Arnold (2013).

is just to have mental states with false contents that are your evidence (whether the state itself or the content doesn't concern me here).

The argument just given closely parallels Williamson's argument from the last section. The objection does as well. The problem lies with the second premise. Although we cite non-factive mental states to specify and evaluate the subject's reasons, it doesn't follow that the non-factive mental states cited are identical to those reasons.

Consider the example from the last section again. The subject believes she is ill because she drank poison. That is how we would specify her reason if we were to believe that she really drank the poison. What if we didn't? Her belief that she is ill might nonetheless be rational. We could give a rationalizing explanation of why she holds that belief by saying that she believes she is ill because she believes she drank poison. The second belief rationalizes the first so long as the second belief is itself rationally held. Even if it is not rationally held, we could still specify what her reason is by saying that she believes she is ill because she believes she drank the poison. In this case her reason fails to do what reasons *qua* reasons are supposed to do: rationalize belief. So, her reason is epistemically defective. But it is still a reason, despite not being a good reason (in either the subjective normative or objective normative sense).[179]

Should we identify her reason with the mental states we cite in specifying and evaluating her reason? No. As we saw in the last section, issues arise when you identify the subject's reason with whatever you cite to specify or evaluate it. We would specify her reason differently if we were to evaluate it differently, but it would still be the same reason.

---

[179] Cf. Turri (2009).

Furthermore, non-factive mental states are ill-suited to explain the role reasons play in objectively favoring belief. Just as reasons *qua* reasons rationalize belief, reasons *qua* reasons objectively favor belief.[180] If your reasons fail to objectively favor the belief that they are reasons for, then they are defective *qua* reasons. They aren't defective in the same way they would be if they failed to rationalize it. However, there is nonetheless something wrong with them *qua* reasons. This point seems to be widely (if obliquely) recognized in the defeasible reasoning literature on the Gettier Problem. The subject of a Gettier case has reasons that rationalize her belief. This is why her true belief is justified. However, there is nonetheless something wrong with her reasons. In virtue of their success in rationalizing the subject's belief, they are subjective normative reasons. However, they are not objective normative reasons. I take it this is why people describe her justifying reasons as "defective" (Audi 1993: 205), "rotten" (de Almeida & Fett 2016) and "not worth having" (Klein 2017).

This further feature of her reasons that makes them objective normative (in the good case) cannot be explained by her non-factive mental states. The Gettier victim and the subject who knows on the basis of the same justifying reasons are identical with respect to their non-factive mental states. Environmental factors distinguish them. The upshot is that the subject's reasons are janus-faced: they are assessed along one dimension by looking in at her mental states and along another dimension by looking out to the world.

Turri (2009) has argued that we can accommodate this point simply in terms of how beliefs are epistemically assessed. Whether a reason is a good reason depends on non-mental

[180] Cf. Dancy (2002), Littlejohn (2012a), Alvarez (2018).

facts, but the reason itself is a mental state. As we have seen, he is right that it is important to distinguish the question of what reasons are from what makes them good reasons. He thinks that reasons are mental states, and these mental states are epistemically assessed in light of the facts.

In assessing his proposal, it is important to bear in mind that to really understand what reasons are, we must be clear about why they are assessed in light of the facts. Why is it that otherwise good reasons can be defeated by facts of which the subject is unaware? It is because they make counterarguments available, which is just to say they make countermoves in the activity of reasoning available. When I believe that someone in my class owns a Ferrari because Mr. Nogot told me he does, my reason is defeated by the fact (of which I am unaware) that Nogot is lying.[181]That is, that fact makes available a counterargument that trumps my reason. As Schroeder (2021: Chapter 2) points out, reasons are competitors. This is not merely a true generalization; that is what makes them reasons. To understand reasons as reasons is to understand them *qua* competitors. To understand them as episodes of reasoning is to do just this. Episodes of reasoning are the episodes they are because of how they relate to potential countermoves. Reasoning is an essentially competitive activity, in the sense that the moves put forth within the activity are essentially subject to challenge. The epistemic assessment of reasons depends on the epistemic assessment of the challenges, and this depends on the facts. This is why the facts matter.

It is fine to say reasons are mental states so long as we are clear that it is true only because we are thinking of mental states insofar as they constitute episodes of a competitive

[181] The example is from Lehrer (1965).

activity (i.e., reasoning). If so, however, we aren't giving a reductive account of reasons in terms of mental states. In fact, we are making sense of mental states in terms of how they constitute episodes of reasoning. This is how I proposed in section (I) we make sense of reflective beliefs.

So, saying that reasons *are* mental states is perhaps a bit misleading. It makes it sound as if we first understand what reasons are and then go on to make sense of their epistemic assessment afterward. But if their epistemic assessment (i.e., the manner in which they compete with one another) is what is essential to them *qua* reasons, then this is not a viable order of explanation. If Turri and company agree with me about this, then my criticism is not that they unambiguously express the wrong view, but rather that they don't unambiguously express the right one.[182]

6.4 Facts and Favoring

Here is an argument that reasons are facts. Reasons objectively favor the beliefs for which they are reasons. Only facts can objectively favor. Therefore, reasons are facts.[183]

The first premise is plausible because it seems that reasons non-accidentally favor the beliefs for which they are reasons. When you provide a reason for a belief and later figure out that you had a justified, false belief or were Gettiered, it seems clear that your reason did less for your belief than you had hoped.[184]You hoped for more than rationalization, you hoped for favoring.

[182] To paraphrase Boyle (2011).
[183] A similar argument is found in Littlejohn (2012a: 147, 228).
[184] Cf. Audi (1993: 205), Steup (1999: Chapter 1), Littlejohn (2012a) de Almeida (2016) and Klein (2017).

Furthermore, the argument goes, only facts can favor. If you infer a belief from a false lemma, then that lemma might rationalize your belief. It nonetheless doesn't favor it.[185][186]

So, it might seem that reasons must be facts. However, this is not quite right. The problem parallels those of the previous two sections. It may well be the case that if the subject's reason for believing favors her belief, we specify her reason by citing a fact. However, it doesn't follow that her reason is identical to that fact.

There are further reasons to resist the identification of epistemic reasons with facts. Suppose I reason from a false lemma. Does it follow that I formed a belief for no reason at all? That would be a doctrinaire conclusion. I may not have had an objective normative reason, but surely, I had a reason! I reasoned, after all. The proponent of factive reasons cannot appeal to the *fact* that I believed the false lemma here. Although that is a fact, it is not a normative reason to believe as I do. However, the argument for factive reasons turns on the claim that only facts can be normative reasons. This is why they say it must be the fact believed rather than the belief itself that is the reason.

Proponents of factive reasons will have to say that although I had no reason, I had an apparent reason.[187] This is a way of mitigating the discomfort of biting the bullet by acknowledging something reason-like that I really did have. Something close to this is true. If I have a rationally held a belief, then I have sufficient subjective normative reason for that belief.

---

[185] This is a recurring motif throughout Littlejohn (2012a, b).
[186] It might be objected that some falsehoods are epistemically benign. See Saunders & Champawat (1964), Hilpinen (1988), Klein (1996: 106, 2008), Hawthorne (2004: 57), Warfield (2005), Coffman (2008), Fitelson (2010, 2017), Feit & Cullison (2011), Arnold (2013), Hiller (2013), de Almeida (2017), Hawthorne & Rabinowitz (2017), Buford & Cloos (2018), Turri (2019), Luzzi (2019), Zhao (2022). I have dealt with this issue in Chapter 3.
[187] Cf. Littlejohn (2012a: 120).

Quite plausibly, subjective normative reasons are apparently (to one in the subject's epistemic position) objective normative reasons. This was the conclusion of chapter 2. So, we can say that the subject had a merely apparent objective normative reason. Although it was merely apparent that her reason was objective normative, it does not follow that it was also merely apparent that she had a reason. That sounds like something that would only be said under philosophical duress. She really (as opposed to merely apparently) engaged in reasoning, so she really did traffic in reasons.

This is made even more clear when we consider reasons that aren't even subjective normative. Consider cases of unjustified belief. Suppose I believe I will win the lottery because I asked a Magic 8 Ball. I shook it and it returned the answer "It is decidedly so." I don't even have a *prima facie* reason for believing I will win. So, I don't have a subjective normative reason and I definitely don't have an objective normative reason. But I still have a reason. I used something as a reason. If you asked me what my reason was, I would tell you it was because the 8 Ball said so. I would be right about that much, despite being wrong about the quality of my reason.

The proponent of factive reasons recognizes that reasons *qua* reasons objectively favor the beliefs for which they are reasons. However, it doesn't follow that every reason succeeds in doing so. Part of what it is to be a reason is to be the kind of thing that gets assessed in light of the facts. When a reason succeeds along this dimension of assessment, it is objective normative. However, some reasons fail along this dimension of assessment. What makes them reasons is that they are the kind of thing that gets assessed along this dimension, not that they are positively assessed in every case. What is the kind of thing that is properly assessed along this dimension? Episodes of reasoning. By taking the Copernican Turn we can understand good reasons and bad

reasons as fundamentally the same type of thing (i.e., episodes of the same activity) while also accounting for how the facts bear differently on their respective assessment.

Furthermore, reasons are competitors and facts as such are not. I think this point is easy to miss, because people often think (I suspect) that the competitive nature of reasons can be explained in terms of probabilistic relations between facts. We should be careful here. Probabilistic relations do not make facts competitors: facts don't mind one another, so to speak. When one fact makes another fact improbable, there is no interesting sense in which the two are competing and the former is winning. Facts are like the marbles on the scale discussed earlier. The marbles weigh different amounts, but that doesn't mean they are opposing each other. We might put them on a scale and bring it about that their independently obtaining weights cause them to exert opposing forces on the scale, but that doesn't mean the marbles are essentially such as to oppose one another. Just as marbles have weights independently of what we do with them, facts stand in relations (e.g., covariation) independently of whether we reason about them. Just as the weights of the marbles enable us to use them to exert opposing forces on the scale, the relations that hold between facts enable us to reason about them. When we do so, we engage in a competitive activity. But the facts that enable us to do so are not themselves competitors. Proponents of factive reasons conflate necessary conditions for reasoning with reasons themselves.

Those trying to make sense of reasons by identifying them with facts are stalking an insight though. It is essential to reasons that they are assessed relative to the facts. The facts form a tribunal. One would fail to understand what reasons are if one weren't aware of the tribunal. Furthermore, the facts are only able to form this tribunal because they stand in appropriate

relations, such as covariation. The Copernican Turn does not deprioritize the facts. The Copernican Turn I recommend is not one where, instead of understanding facts first and reasons second, we understand reasons first and facts second. As I have said above, we only understand reasons in terms of how they are assessed, and this makes essential reference to the facts. So, instead of replacing one reductive analysis with another that runs in the opposite direction, I recommend replacing reductive analyses altogether with an account that situates reasons, facts, and mental states in a common network. The activity of reasoning is that network. We understand the epistemic significance of facts and mental states in terms of how they bear on the form of assessment proprietary to the activity, and we understand the activity in terms of a form of assessment that makes essential reference to facts and mental states. The account is nonetheless "Copernican" in the sense that it invites us to abandon the most obvious strategy and instead go in the opposite direction. Instead of explaining reasons in terms of a prior understanding of something else, we instead explain the rational significance of everything (facts and mental states) in terms of how they bear on the kind of assessment that makes reasons intelligible in the first place.

6.5 Amphibious Reasons

In this section I want to briefly discuss another family of positions on the metaphysics of reasons. Unlike in the previous sections, I won't try to extract a syllogism from the literature. Instead, I will show that one of the intuitions motivating the position that reasons are amphibious can be accounted for in light of the above. I want to show that we can vindicate an insight to which proponents of the amphibious view have drawn our attention without saying that reasons themselves are amphibious (at least not in the way proponents of this position typically think). Rather, their assessment is amphibious.

Reasons are amphibious just in case they live, so to speak, both in the mind and in the world. What kind of creature could live in both environments? Perhaps true mental states.[188] Their truth is worldly, but their mental status is not. Perhaps reasons are a subset of veridical mental states: known facts.[189] Perhaps reasons are facts that are manifest to you[190] or that you are in a position to know.[191]

It is impossible to adequately assess any of these views independently of the research programs (e.g., knowledge-first epistemology, disjunctivist responses to skepticism, etc.) in which they make their appearance. However, the last three sections put us in a position to vindicate an insight they are stalking: reasons are assessed both in light of the subject's mental states and the facts. The epistemic assessment of reasons is janus-faced. It is epistemically appropriate to assess the subject's reasons both relative to her mental states and relative to the facts. Janus-faced assessment is essential to reasons. Something not appropriately assessed in this way is not a reason. On account of being subject to this kind of epistemic assessment, reasons have a foothold both in the mind and the world.

The amphibious nature of reasons can be understood in terms of their amphibious epistemic assessment. This means that we can preserve the amphibian insight without also trying to find something other than reasoning and its epistemic assessment that has both mental and non-mental aspects. Certain cognitive relations, such as the relation of manifestation or the relation of being in a position to know, hold between a subject and an object of knowledge. The

[188] See Mitova (2015).
[189] See Williamson (2000).
[190] See Neta (2008; 2018), cf. McDowell (1994); Neta & Pritchard (2007), Pritchard (2012).
[191] See Gibbons (2013) and Simion (2023).

former is a mental subject and the latter is non-mental (typically). Recognition of the fact that reasons straddle the mental/non-mental gap can incline us to think that reasons must be identical to a relation that has relata on both sides. However, I hope to have shown that we needn't identify reasons with any such thing in order to account for the fact that reasons don't fit squarely on either side.

A similar point applies to pluralism about epistemic reasons. Noticing the diversity of roles epistemic reasons play, some people have expressed doubts that anything in our ontology can play all these roles.[192] In a way, they are right. There is nothing we can pick out under a guise that doesn't make reference to reasoning itself that can do this. However, we can be monists about reasons and incorporate the pluralist insight that several ontologically distinct things (e.g., mental states, propositions and facts) bear on the epistemic assessment of reasons. The Copernican Turn enables this by inverting the order of explanation. Instead of identifying reasons with everything we cite while reasoning, we should instead explain the epistemic significance of the things we cite in terms of how they bear on the kind of epistemic assessment that constitutes the activity of reasoning.

6.6 Conclusion

I have shown that many arguments made in defense of claims about what reasons are really have more to do with how reasons are epistemically assessed. To figure out what reasons are, we just need to ask what is appropriately assessed in that way. Episodes of reasoning are. So, reasons are episodes of reasoning. This invites a Copernican Turn in the metaphysics of epistemic reasons. Rather than trying to make sense of reasons by identifying them with

[192] See Kelly (2008), Rysiew (2011; 2018) and Fogal (2018).

something antecedently understood, such as facts or mental states, we should understand the epistemic significance of these things in terms of how they bear on the epistemic assessment of episodes of reasoning.

Bibliography:

Alfano, M., & Levy, N. (2019). Knowledge from vice: deeply social epistemology. *Mind*, *129*(515), 887–915. https://doi.org/10.1093/mind/fzz017

Alston, W. (1993). Epistemic desiderata. *Philosophy and Phenomenological Research*, *53*(3), 527–551. https://doi.org/10.2307/2108079

Alvarez, M. (2018). Reasons for action, acting for reasons, and rationality'. *Synthese*, *195* (8),3293–3310.

Anscombe, G.E.M. (1957). *Intention*. Harvard University Press.

Armstrong, D.M. (1973). *Belief, truth and knowledge*. Cambridge University Press.

Arnold, A. (2011). Some Evidence is False. *Australasian Journal of Philosophy*, *91*(1), 165–172. https://doi.org/10.1080/00048402.2011.637937

Ashley, J. & Tomasello, M. (1998). Cooperative problem-solving and teaching in pre-schoolers. *Social Development 17*, 143–63. https://doi.org/10.1111/1467-9507.00059

Audi, R. (1993) *The structure of justification*. Cambridge University Press.

Audi, R. (1994). Dispositional beliefs and dispositions to believe". *Noûs*, *28*(4), 419–34. https://doi.org/10.2307/2215473

Audi, R. (2003) *Epistemology: A contemporary introduction to the theory of knowledge* (2nd Ed.). Routledge.

Baars, B. J. (1986). *The Cognitive Revolution in Psychology*. Guilford Press.

Bach, Kent (1985). A rationale for reliabilism. *The Monist 68*, 246–63. https://doi.org/10.5840/monist198568224

Baehr, J. (2006). Character, reliability, and virtue epistemology. *The Philosophical Quarterly*, *56* (223), 193–212. https://doi.org/10.1111/j.1467-9213.2006.00437.x

Baier, A. (1981). Cartesian persons. *Philosophia, 10,*169-188. https://doi.org/10.1007/BF02380762

Bakeman, R. & Adamson, L. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interactions. *Child Development, 55,* 1278–89. https://doi.org/10.2307/1129997

Baker-Hytch, M. & Benton, M. (2015). Defeatism defeated. *Philosophical Perspectives*, *29*(1), 40-66. https://doi.org/10.1111/phpe.12056

Ball, B. & Blome-Tillman, M. (2014). Counter closure and knowledge despite falsehood. *Philosophical Quarterly,* 64(257), 552-68. https://doi.org/10.1093/pq/pqu033

Ballantyne, N. (2015). The significance of unpossessed evidence. *The Philosophical Quarterly, 65*(260), 315-335. https://doi.org/10.1093/pq/pqu096

Barker, J. (1976). What you don't know won't hurt you? *American Philosophical Quarterly*, *13,* 303-8.

Barkow, J., Cosmides, L. & Tooby, J. (1992). *The adapted mind: Evolutionary psychology and the generation of culture*. Oxford University Press.

Barsalou, L.W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 588-609. DOI: 10.1017/s0140525x99002149

Barsalou, L.W. (2005). Continuity of the conceptual system across species*. Trends in Cognitive Sciences*, *9*, 309-11. https://doi.org/10.1016/j.tics.2005.05.003

Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617-45. DOI: 10.1146/annurev.psych.59.103006.093639

Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy*. Academic Press.

Bergmann, M. (2006) *Justification without awareness: A defense of epistemic externalism*. Oxford University Press.

Bermudez, J. (2003). *Thinking without words*. Oxford University Press.

Billig, M. (1996) *Arguing and thinking: A rhetorical approach to social psychology*. Cambridge University Press.

Bird, A. (2007). Justified judging. *Philosophy and Phenomenological Research,* 74 (1), 81–110. https://doi.org/10.1111/j.1933-1592.2007.00004.x

Blum-Kulka, S., Blondheim, M. & Hacohen, G. (2002). Traditions of dispute: From negotiations of talmudic texts to the arena of political discourse in the media. *Journal of Pragmatics*, *34*(10 –11),1569– 94. https://doi.org/10.1016/S0378-2166(02)00076-0

Bogardus, T. (2014). Knowledge under threat. *Philosophy and Phenomenological Research*, *88*(2), 289-313. https://doi.org/10.1111/j.1933-1592.2011.00564.x

BonJour, L. (1985). *The structure of empirical knowledge*. Harvard University Press.

BonJour, L. (2003). In BonJour, L. & Sosa, E. (authors) *Epistemic justification*: *internalism vs. externalism, foundations vs. virtues.* Oxford University Press.

BonJour, L. (2010). *Epistemology: Classic problems and contemporary responses* (2nd Edition). Rowman & Littlefield Publishing.

Borges, R. (2017) Inferential knowledge and the gettier conjecture. In. R. Borges, C. de Almeida, & P. Klein (eds.), *Explaining Knowledge: New Essays on the Gettier Problem* (pp. 292-311). Oxford University Press.

Boyle, M. (2011). Making up your mind and the activity of reason. *Philosophers' Imprint 11*: 1-24.

Boyle, M. (2012). Essentially rational animals. In G. Abel & J. Conant (Eds.), *Rethinking Epistemology*: *Volume 2* (pp. 395-428). De Gruyter. https://doi.org/10.1515/9783110277944.395

Brandom, R. (1994). *Making it explicit*. Harvard University Press.

Brewer, B. (1999). *Perception and reason*. Oxford: Oxford University Press.

Broome, J. (2004). Reasons. In R.J. Wallace (ed.) *Reason and Value: Themes from the Philosophy of Joseph Raz* (pp.204-228). Oxford University Press.

Brown, J. & Simion, M. (2021) *Reasons, justification and defeat*. Oxford University Press.

Brueckner, Anthony. (2009). E=k and perceptual knowledge. In P. Greenough & D. Pritchard (eds.). *Williamson on Knowledge* (pp. 5-12). Oxford University Press.

Bruner, J. (1972). The nature and uses of immaturity. *American Psychologist*, *27*, 687-708.

Buckner, C. (2018). Empiricism without magic: transformational abstraction in deep convolutional neural networks. *Synthese*, *195*, 5339-5372. https://doi.org/10.1007/s11229-018-01949-1

Buford, C. & Cloos, C.M. (2018). A dilemma for the knowledge despite falsehood strategy. *Episteme*, *15*(2), 166-82. doi:10.1017/epi.2016.53

Burge, T. (2010). *Origins of objectivity*. Oxford University Press.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10,* 425-55. https://doi.org/10.1080/01690969508407111

Bybee, J. Sequentiality as the basis of constituent structure. In T. Givon, & B. Malle (eds.) *From pre-language to language* (pp. 313-35). Benjamins.

Byrne, R. W. & Whiten, A., eds. (1988). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press.

Call, J. (2010). Do apes know that they could be wrong? *Animal Cognition*, *13*, 689 –700. DOI: 10.1007/s10071-010-0317-x

Call, J. & Tomasello, M. (2007). *The gestural communication of apes and monkeys.* Lawrence Erlbaum.

Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence". *Philosophy and Phenomenological Research*, *78*(2), 275-311. https://doi.org/10.1111/j.1933-1592.2009.00245.x

Carruthers, P. (2006). *The architecture of the mind: Massive modularity and the flexibility of thought.* Oxford University Press.

Carter, J.A. (2014). Robust virtue epistemology as anti-luck epistemology: A new solution. *Pacific Philosophical Quarterly*, *97*, 140-55. https://doi.org/10.1111/papq.12040

Carter, J.A. (2020). Intellectual autonomy, epistemic dependence and cognitive enhancement. *Synthese, 197,* 2937-2961. https://doi.org/10.1007/s11229-017-1549-y

Carter, J.A. (2022a). Epistemic autonomy & externalism. In J. Matheson & K. Lougheed (eds.) *Routledge studies in epistemology: Epistemic autonomy* (pp. 21-41). Routledge.

Carter, J.A. (2022b). *Autonomous knowledge: radical enhancement, autonomy and the future of knowing*. Oxford University Press.

Carter, J. A. & McKenna, R. (2019). Sosa versus kornblith on grades of knowledge". *Synthese. 196*, 4989-5007. DOI: https://doi.org/10.1007/s11229-018-1689-8

Cheney, D.L. & Seyfarth, R.M. (1980). Vocal recognition in free ranging vervet monkeys. *Animal Behavior*, *28*, 362-7. https://doi.org/10.1016/S0003-3472(80)80044-3

Chomsky, N. (1959). Review of B. F. Skinner's *Verbal Behavior*. *Language, 35*, 26–58.

Clark, M. (1963) "Knowledge and Grounds: A Comment on Mr. Gettier's Paper". *Analysis, 24* (2),46-8.

Clark, H. (1996). *Uses of Language*. Cambridge University Press.

Coffman, E.J. (2008). Warrant Without Truth. *Synthese 162*(2), 173-94. https://doi.org/10.1007/s11229-007-9178-5

Coffman, E.J. (2010) Misleading dispositions and the value of knowledge. *Journal of Philosophical Research*, *35*, 241-58. DOI: 10.5840/jpr_2010_13

Corkum, V. & Moore, C. (1995). Development of joint visual attention in infants. In C. Moore & P.J. Dunham (eds.) *Joint attention: Its origins and role in development* (pp. 61-83). Erlbaum.

Cohen, S. (1984). Justification & truth. *Philosophical Studies*, *46* (3), 279-95. DOI 10.1007/bf00372907

Comesaña, J. (2005). Unsafe knowledge. *Synthese, 146,* 395–404. DOI: https://doi.org/10.1007/s11229-004-6213-7

Conee, E. (2007). Disjunctivism and anti-skepticism. *Philosophical Issues*, *17* (1), 16–36. https://doi.org/10.1111/j.1533-6077.2007.00120.x

Conee, E. & Feldman, R. (2001). Internalism defended. In H. Kornblith (ed.), *Internalism and Externalism in Epistemology* (pp. 1-18). Blackwell.

Conee, E. & Feldman, R. (2004). *Evidentialism: Essays in epistemology.* Oxford University Press.

Conee, E. & Feldman, R. (2008). Evidence. In Q. Smith (ed.) *Epistemology: New essays*. (pp. 83-104). Oxford University Press.

Conee, E. & Feldman, R. (2011). Replies. In T. Dougherty (ed.) *Evidentialism & its discontents*. Oxford University Press.

Croft, W. (2001). *Radical construction grammar: Syntactic theory in typological perspective.* Oxford University Press.

Danchin, E., & Luc-Alain, G. (2004). Public information: From nosy neighbors to cultural evolution. *Science*, *305*, 487–91. DOI: 10.1126/science.1098254

Dancy, J. (1995). Why there is really no such thing as a theory of motivation. *Proceedings of the Aristoteleian Society*, *95*, 1-18. 10.1093/aristotelian/95.1.1

Dancy, J. (2002). *Practical reality*. Oxford University Press.

Davidson, D. (1980). *Inquiries into truth and interpretation*. Oxford University Press.

Davidson M.C., Amso D., Anderson L.C., Diamond A. (2006). Development of cognitive control and executive functions from 4–13 Years: Evidence from manipulations of memory, inhibition, and task switching". *Neuropsychologia, 44,* 2037–78. DOI: 10.1016/j.neuropsychologia.2006.02.006

de Almeida, C. (2017). Knowledge, benign falsehoods and the gettier problem. In. R. Borges, C. de Almeida, & P. Klein (eds.) *Explaining knowledge: New essays on the gettier problem* (pp. 292-311). Oxford University Press.

de Almeida, C. & Fett, J.R. (2016). Defeasibility & gettierization: A reminder. *Australasian Journal of Philosophy*, *94*(1), 152-69. DOI: 10.1080/00048402.2015.1009127

Dessalles, J.-L. (2007). *Why we talk: The evolutionary origins of language*. Oxford University Press.

Descartes, R. (2009). *Selected philosophical writings* (J. Cottingham, R. Stoothoff & D. Murdoch, Trans.). Cambridge University Press.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-68. DOI: 10.1146/annurev-psych-113011-143750.

Dogramaci, S. (2013). Communist conventions for deductive reasoning. *Noûs*, *49*(4), 776-799. https://doi.org/10.1111/nous.12025

Dogramaci, S. (2020). What is the function of reasoning? On Mercier & Sperber's argumentative and justificatory theories. *Episteme*, 17(3), 316-330.

Dougherty, T. (2011). In defense of propositionalism about evidence. In T. Dougherty. (ed.) *Evidentialism & its discontents* (pp. 226-35). Oxford University Press.

Dummett, M. (1983). Language and truth. In R. Harris (ed.) *Approaches to language* (pp. 95-125). Pergamon Press.

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology, 6*, 178– 90. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R. I. M. & Shultz, S. (2003). Evolution of the social brain. *Science.2*, 1160 – 61. DOI: 10.1126/science.1145463

Engel, M. (1992). Is epistemic luck compatible with knowledge? *The Southern Journal of Philosophy*, 30 (2), 59-75. https://doi.org/10.1111/j.2041-6962.1992.tb01715.x

Evans, Gareth. (1982) *The varieties of reference*. J. McDowell (ed.). Clarendon Press.

Evans, J. St. B. T. (2002). Logic and human reasoning: An assessment of the deduction paradigm. *Psychological Bulletin*, *128*(6), 978 –96.

Feit, N. & Cullison, A. (2011). When does falsehood preclude knowledge? *Pacific Philosophical Quarterly*, *92*, 283–304. https://doi.org/10.1111/j.1468-0114.2011.01397.x

Feldman, R. (1988). Subjective and objective justification in ethics and epistemology. *The Monist*, *71*(3), 405–19. https://doi.org/10.5840/monist198871326

Feldman, R. (2003). *Epistemology.* Pearson.

Fitelson, B. (2010). Strengthening the case for knowledge from false belief. *Analysis*, *70*(4), 666-9. https://doi.org/10.1093/analys/anq078

Fitelson, B. (2017). Closure, counter-closure and inferential knowledge. In R. Borges, C. de Almeida., & P. Klein. (eds.) *Explaining knowledge: New essays on the gettier problem* (pp. 312-24). Oxford University Press.

Fodor, J. (1975). *The language of thought*. Harvard University Press.

Fodor, J. (1983). *The modularity of mind*. MIT Press.

Fodor, J. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind.* MIT Press.

Fogal, D. (2018). Deflationary pluralism about motivating reasons. In V. Mitova. (ed.) *The factive turn in epistemology* (pp.193-218). Oxford University Press.

Fogelin, R. (1994). *Pyrrhonian reflection on knowledge and justification*. Oxford University Press.

Foley, R. (1993). *Working without a net: A study in egocentric epistemology*. Oxford University Press.

Foley, R. (2012). *When is true belief knowledge?* Princeton University Press.

Foley, M. & Ratner, H. (1997) Children's encoding in memory for collaboration: A way of learning from others. *Cognitive Development*, *13*, 91-108.

Frawley, W. (1997). *Vygotsky & C=cognitive science: Language and the unification of the social and computational mind*. Harvard University Press.

Fumerton, R. (1995). *Metaepistemology & skepticism*. Rowman & Littlefield.

Garon N., Bryson S.E., Smith I.M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134,* 31–60. DOI: 10.1037/0033-2909.134.1.31

Gauker, C. (2011). *Words and images: An essay on the origin of ideas*. Oxford University Press.

Gettier, E. (1963). Is justified true belief knowledge? *Analysis*, *23*, 121-3.

Ginsborg, H. (2006). Reasons for belief. *Philosophy and Phenomenological Research*, *72* (2), 286-318. https://doi.org/10.1111/j.1933-1592.2006.tb00562.x

Gibbons, J. (2010). Things that make things reasonable. *Philosophy & Phenomenological Research*, *81*(2), 335-61. https://doi.org/10.1111/j.1933-1592.2010.00373.x

Gibbons, J. (2013). *The norm of belief.* Oxford University Press.

Godfrey-Smth, P. & Yegnashanakaran, K. (2011). Reasoning as deliberative in function but dialogic in dtructure and origin. *Behavioral and Brain Sciences*, *34*(2), 80. https://doi.org/10.1017/S0140525X10002906

Goldberg, S. (2015). Epistemic entitlement and luck. *Philosophy and Phenomenological Research*, *91*(2), 273-302. https://doi.org/10.1111/phpr.12083

Goldberg, S. (2018) *To the best of our knowledge: Social expectations and epistemic normativity*. Oxford University Press.

Goldberg, S. (2019). A normative account of epistemic luck. *Philosophical Issues*, 29(1), 97-109. https://doi.org/10.1111/phis.12143

Goldberg, S. (2020) *Conversational pressure*. Oxford University Press.

Goldberg, S. & Matheson, J. (2020). The impossibility of mere animal knowledge for reflective subjects. *Erkenntnis*, *85*(4): 829-40. https://doi.org/10.1007/s10670-018-0052-2

Goldman, A. (1976). Discrimination and perceptual knowledge. *The Journal of Philosophy*, *73(*20), 771-91. DOI: 10.2307/2025679

Goldman, Alvin. (2009). Williamson on knowledge and Evidence. In P. Greenough & D. Pritchard (eds.) *Williamson on Knowledge.* (pp. 73-92). Oxford University Press.

Goodman, S. (1981). The integration of verbal and motor behavior in preschool children. *Child Development*, *52*, 280-9. https://doi.org/10.2307/1129241

Graham, P. (2000). Transferring knowledge. *Noûs*, *34*(1), 131-52. DOI: 10.1111/0029-4624.00205

Graham, Peter J. (2012). Epistemic Entitlement. *Noûs*, *46*(3), 449-482.

Greco, J. (2003). Knowledge as credit for true belief. In M. DePaul & L. Zagzebski. (eds) *Intellectual Virtue: Perspectives from Ethics and Epistemology* (pp. 111-34). Clarendon Press.

Greco, J. (2010). Achieving knowledge: A virtue theoretic account of epistemic normativity. Cambridge University Press.

Greco, J. (2012). A (different) virtue epistemology. *Philosophy and Phenomenological Research*, *85*(1), 1–26. https://doi.org/10.1111/j.1933-1592.2011.00567.x

Greenwood, J. D. (1999). Understanding the 'cognitive revolution' in psychology. *Journal of the History of the Behavioral Sciences*, *35*(1), 1–22. https://doi.org/10.1002/(SICI)1520-6696(199924)35:1<1::AID-JHBS1>3.0.CO;2-4

Griffin, D.R. (1978). Prospects for a cognitive ethology. *Behavioral and Brain Sciences, 4*, 527–38. https://doi.org/10.1017/S0140525X00076524

Hagler, D. A. & Brem, S. K. (2008). Reaching agreement: The structure & pragmatics of critical care nurses' informal argument. *Contemporary Educational Psychology, 33*(3), 403–24. https://doi.org/10.1016/j.cedpsych.2008.05.002

Hahn, U. & Oaksford, M. (2007). The rationality of informal argumentation: A bayesian approach to reasoning fallacies. *Psychological Review*, *114*,704 –32. https://doi.org/10.1037/0033-295X.114.3.704

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences of the United States of America*, *98*, 5359–5362. 10.1073/pnas.071600998

Hanks, P. (2015). *Propositional content*. Oxford University Press.

Hare, B., Call, J. & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour, 68*, 571–81. https://doi.org/10.1006/anbe.1999.1377

Harman, G. (1973). *Thought*. Princeton University Press.

Harman, G. (1980). Reasoning and explanatory coherence. *American Philosophical Quarterly 17*(2), 151–7.

Hawthorne, John. (2004) *Knowledge and Lotteries*. Oxford University Press.

Hawthorne, J. & Gendler, T. (2005). Real guide to fake barns: A catalogue of gifts for your epistemic enemies. *Philosophical Studies*, *124*(3), 331–52. https://doi.org/10.1007/s11098-005-7779-8

Hawthorne, J. & Rabinowitz, D. (2017). Knowledge and False Belief. In R. Borges, C. de Almeida, & P. Klein (eds.) *Explaining knowledge: New essays on the gettier problem* (pp.325–45). Oxford University Press.

Henrich, J. & Boyd, R. (2002). On modeling cognition and culture: Why cultural evolution does not require replication of representations. *Journal of Cognition and Culture, 2*, 87–112. https://doi.org/10.1163/156853702320281836

Hetherington, S. (1999) Knowing failably. *Journal of Philosophy 96,* 565 587. DOI: 10.2307/2564624

Hieronymi, P. (2021). Reasoning first. In R. Chang & K. Sylvan (eds.) *Routledge Handbook of Practical Reasoning* (pp. 349–65). Routledge Press.

Hiller, A. (2013). Knowledge essentially based upon false belief. *Logos & Episteme, 4*(1),7–19. 10.5840/logos-episteme20134136

Hiller, A. & Neta, R. (2007). Safety and luck. *Synthese, 158*(3), 303–13. 10.1007/s11229-006-9041-0

Hilpinen, R. (1971). Knowledge & justification. *Ajatus, 33*,7–39.

Hilpinen, R. (1988). Knowledge and conditionals. In J. Tomberlin. (ed.) *Philosophical Perspectives Volume 2*, *Epistemology* (pp. 157–82). Ridgeview.

Horty, J. (2007). Reasons as defaults. *Philosophers' Imprint*, *7,* 1–28.

Horty, J. (2012). *Reasons as defaults*. Oxford University Press.

Howard-Snyder, D., Howard-Snyder, F. & Feit, N. (2003). Infallibilism and gettier's legacy. *Philosophy & Phenomenological Research, 66*(2), 304–27. https://doi.org/10.1111/j.1933-1592.2003.tb00263.x

Hrdy, S. B. (2009). *Mothers and others*. Belknap Press.

Huemer, M. (2006). Phenomenal conservatism and the internalist intuition. *American Philosophical Quarterly*, *43(*2), 147–158.

Huemer, M. (2007). Compassionate Phenomenal Conservatism. *Philosophy and Phenomenological Research*, *74*(1), 30–55. DOI: 10.1111/j.1933-1592.2007.00002.x

Humphrey, N. K. (1976). The social function of intellect. In P. P. G. Bateson & R. A. Hinde (eds.) *Growing points in ethology* (pp. 303–17). Cambridge University Press.

Hurley, S. (2003). Animal action in the space of reasons. *Mind & Language, 18(*3),231–56. DOI: 10.1111/1468-0017.00223

Jackendoff, R. S. (1996). How language helps us think. *Pragmatics and Cognition*, *4*(1), 1–34. 10.1075/pc.4.1.03jac

Jenkin, Z. (2022). Perceptual learning and reasons-responsiveness. *Noûs, 57*(2), 481–508.

Johnsen, B. (1974). Knowledge. *Philosophical Studies*, *25*(4), 273–82.

Kaplan, M. (1985). It's not what you know that counts. *Journal of Philosophy*, 82(7), 350–63.

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

Kelly, T. (2008). Evidence: Fundamental concepts and the phenomenal conception. *Philosophy Compass, 3*(5), 933–955.

Kelp, C. (2009). Knowledge and safety. *Journal of Philosophical Research*, *34*, 21–31. 10.5840/jpr_2009_1

Kelp, C. (2013). Knowledge: The safe-apt view. *Australasian Journal of Philosophy*, *91*(2), 265–78. 10.1080/00048402.2012.673726

Kelp, C. (2016). Justified belief: Knowledge first-style. *Philosophy and Phenomenological Research*, *91*(1), 79–100. 10.1111/phpr.12272

Kelp, C. (2018). *Good thinking: A knowledge first virtue epistemology*. Routledge.

Kelp, Christoph. (2021). Theory of inquiry. *Philosophy & Phenomenological Research*, *103*(2), 359–84. 10.1111/phpr.12719

Khalifa, K., & Goldberg, S. C. (2022). Socio-functional foundations in science: The case of measurement. *Philosophical Issues*, *32*, 382–397. 10.1111/phis.12218

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLoS Computational Biology*, *10*(11), https://doi.org/10.1371/journal.pcbi.1003915

Kirkham, R. (1984). Does the gettier problem rest on a mistake? *Mind*, *93*(372), 501–13.

Klein, P. (1971). A proposed definition of propositional knowledge. *Journal of Philosophy*, *68*(16), 471–82.

Klein, P. (1976). Knowledge, causality, and defeasibility. *Journal of Philosophy*, *73*(20):792–812.

Klein, P. (1980). Misleading evidence and the restoration of justification. *Philosophical Studies*, *37*(1), 81–9.

Klein, P. (1996) Warrant, proper function, reliabilism, and defeasibility. In J. Kvanvig (Ed.), *Warrant in contemporary epistemology: Essays in honor of Plantinga's theory of knowledge* (pp. 97–130). Rowman & Littlefield.

Klein, P. (2008) "Useful False Beliefs". In Q. Smith. (Ed.) *Epistemology: New essays* (pp.25–62). Oxford University Press.

Klein, P. (2017). The nature of knowledge. In R. Borges., C. de Almeida & P. Klein (eds.) *Explaining knowledge: New essays on the gettier problem* (pp. 35–56). Oxford University Press.

Kornblith, H. (2009). Sosa in perspective. *Philosophical Studies, 144*(1), 127-–136. 10.1007/s11098-009-9377-7

Kornblith, H. (2012) *On reflection*. Oxford University Press.

Kornell, N., Son, L., & Terrace, H. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*, *18*, 64 –71. DOI: 10.1111/j.1467-9280.2007.01850.x

Korsgard, C. (1996). *The sources of normativity*. Harvard University Press.

Kripke, S. (2011). *Philosophical troubles, volume 1*. Oxford University Press.

Kvanvig, J.L. (1984). Subjective Justification. *Mind, 93*(369), 71-84. https://doi.org/10.1093/mind/XCIII.369.71

Kvanvig, J.L. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge University Press.

Kvanvig, J.L. (2017). Lessons from Gettier. In R. Borges, C. de Almeida, & P. Klein, *Explaining knowledge: New essays on the gettier problem* (pp. 152–69). Oxford University Press.

Lackey, J. (2005). Testimony and the infant/child objection. *Philosophical Studies, 126*(2), 163–190. https://doi.org/10.1007/s11098-004-7798-x

Lackey, J. (2006). Pritchard's epistemic luck. *Philosophical Quarterly, 56*(223), 284–289. https://doi.org/10.1111/j.1467-9213.2006.00443.x

Lackey, J. (2009). Knowledge and credit. *Philosophical Studies*, *142*(1), 27–42. DOI: 10.1007/s11098-008-9304-3

Lakatos, I. (1978). *The methodology of scientific research programmes, volume 1: Philosophical papers*. Cambridge University Press.

Laland, K. & Hoppitt, W. (2003). Do animals have culture? *Evolutionary Anthropology*, *12*, 150–9. https://doi.org/10.1002/evan.10111

Langacker, R. *Foundations of cognitive grammar, volume 1*. Stanford University Press.

Langacker, R. A usage based model. In M. Barlow & S. Kemmerer, *Usage based models of language* (pp.127–61). SLI Publications.

Lasonen-Aarnio, M. (2010a). Is there a viable account of well-founded Belief. *Erkenntnis*, *72*(2), 205–31. DOI: 10.1007/s10670-009-9200-z

Lasonen, Aarnio, M. (2010b). Unreasonable knowledge. *Philosophical Perspectives, 24*(1), 1–21. DOI: 10.1111/j.1520-8583.2010.00183.x

Lassonen-Aarnio, M. (2014). Higher-order evidence and the limits of defeat. *Philosophy and Phenomenological Research*, *88*(2), 314–345. DOI: 10.1111/phpr.12090

Lehrer, K. (1965). Knowledge, truth and evidence. *Analysis*, *25*, 168–75.

Lehrer, K. (1970). The fourth condition for knowledge: A defense. *Review of Metaphysics*, *24*, 122–128.

Lehrer, K. (1974). *Knowledge.* Clarendon Press.

Lehrer, K. (1990). *Theory of Knowledge*. Routledge.

Lehrer, K. (2017). Defeasible reasoning and representation: The lesson of Gettier. In R. Borges, C. de Almeida & P. Klein (eds) *Explaining knowledge: New essays on the gettier problem* (pp.169-79). Oxford University Press.

Lehrer, K. & Paxson, T. (1968). Knowledge: Undefeated justified true belief. *Journal of Philosophy*, *66*(8), 225-37.

Lewis, D. (1969). *Convention*. Harvard University Press.

Lewis, D. (1973). *Counterfactuals*. Harvard University Press.

Littlejohn, C. (2012a). *Justification and the truth-connection*. Cambridge University Press.

Littlejohn, C. (2012b) No evidence is false. *Acta Analytica*, *28*(2), 145-59. DOI: 10.1007/s12136-012-0167-z

Lord, E. & Maguire, B. (2016). *Weighing reasons*. Oxford University Press.

Lorenz, K. (1965). *Evolution and the modification of behavior*. University of Chicago Press.

Luria, A. (1961). *The role of speech in normal and abnormal behavior*. Boni & Liveright.

Luzzi, F. (2010). Counter-Closure. *Australasian Journal of Philosophy*, *88*(4), 673-83.

Luzzi, F. (2019) *Knowledge from non-knowledge: Inference, testimony and memory*. Cambridge University Press.

Lycan, W. (2006). On the gettier problem problem. In S. Hetherington. (ed.) *Epistemology Futures* (pp. 148-68). Oxford University Press.

MacFarlane, J. (2014). *Assessment-sensitivity: Relative truth and its applications*. Oxford University Press.

Manrique, H. M., Gross, A.N.-M., & Call, J. (2010). Great apes select tools on the basis of their rigidity. *Journal of Experimental Psychology: Animal Behavior Processes, 36*(4), 409–422. https://doi.org/10.1037/a0019296

Marr, D. (1982). *Vision*. Freeman & Company.

McAllister, B. (2024). *Seemings and the foundations of justification: A defense of phenomenal conservatism*. Routledge.

McClelland, J. & Chihara, C. (1985). The surprise examination paradox. *Journal of Philosophical Logic*, *4*(1), 71-89.

McDowell, J. (1994). *Mind and World*. Harvard University Press.

McHugh, C. & Way, J. (2018a). What is Good Reasoning? *Philosophy and Phenomenological Research*, 96(1), 153–174. https://doi.org/10.1111/phpr.12299

McHugh, C. & Way, J. (2018b). What is reasoning? *Mind, 127*(505), 167-196. https://doi.org/10.1093/mind/fzw068

Melis, A., Call, J. & Tomasello, M. (2006). Chimapnzees conceal visual and auditory information from others. *Journal of Comparative Psychology*, *120*, 154-62. DOI: 10.1037/0735-7036.120.2.154

Mercier, H. & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, *34*(2), 57-74. doi: 10.1017/S0140525X10000968

Mikkelsen, L. (2004). *Specifying who.* UC Santa Cruz Dissertation.

Millar, A. (2019). *Knowing by perceiving*. Oxford University Press.

Miller, G. A. (2003). The cognitive revolution: A historical perspective. *Trends in the Cognitive Sciences*, *7*(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9.

Millikan, R. (1984). *Language, thought and other biological categories.* MIT Press.

Miracchi, L. (2015). Competence to know. *Philosophical Studies*, *172*, 29-56. https://doi.org/10.1007/s11098-014-0325-9

Mitova, V. (2015). Truthy psychologism about evidence. *Philosophical Studies*, *172*(4), 1105–1126.

Mitova, V. (2018). Introduction. In V. Mitova (ed.) *The factive turn in epistemology* (pp.1-13). Cambridge University Press.

Moffett, M. (2005). Constructing attitudes. *ProtoSociology, 21*, 105-128. DOI: 10.5840/protosociology2005216

Montminy, M. (2014). Knowledge despite falsehood. *Canadian Journal of Philosophy*, *44*(3-4). DOI: 10.1080/00455091.2014.982354

Moon, A. (2018). How to use cognitive faculties you never knew you had. *Pacific Philosophical Quarterly, 99*(S1):251-275. DOI: 10.1111/papq.12234

Moran, R. (2001). *Authority & estrangement: An essay on self-knowledge*. Princeton University Press.

Moretti, L. & Piazza, T. (2018). Defeaters in current epistemology: Introduction to the special issue. *Synthese, 195*(7), 2845-54. DOI: 10.1007/s11229-017-1551-4

Moro, A. *The Raising of Predicates*. Cambridge University Press.

Moser, P.K. (1988). Review of John Pollock's *contemporary theories of knowledge*. *British Journal for the Philosophy of Science*, *39*(1), 131-40.

Moser, P.K. (1989) *Knowledge & evidence*. Cambridge University Press.

Munroe, W. (2021). Why are you talking to yourself? The epistemic role of inner speech in reasoning". *Noûs,* 56(4), 841-66. DOI: 10.1111/nous.12385

Neta, R. (2008). What evidence do you have? *British Journal for the Philosophy of Science*, *59*(1), 89-119. DOI: 10.1093/bjps/axn003

Neta, R. (2018). "Your evidence is the set of facts that are manifest to you". In V. Mitova (ed.) *The factive turn in epistemology* (pp. 32-49). Cambridge University Press.

Neta, R. & Pritchard, D. (2007). McDowell & the new evil genius. *Philosophy & Phenomenological Research*, *74*(2), 381-96. DOI: 10.1111/j.1933-1592.2007.00023.x

Neta, R. & Rohrbaugh, G. (2004). Luminosity and the safety of knowledge" *Pacific Philosophical Quarterly*, *85*(4), 396-406. 10.1111/j.1468-0114.2004.00207.x

Neuman, Y. (2003.) Go ahead, prove that god does not exist! On high school students' ability to deal with fallacious arguments. *Learning and Instruction*, *13*(4), 367 – 80. https://doi.org/10.1016/S0959-4752(02)00011-7

Neuman, Y., Weinstock, M. P. & Glasner, A. (2006). The effect of contextual factors on the judgment of informal reasoning fallacies. *Quarterly Journal of Experimental Psychology, Section A: Human Experimental Psychology*, *59*, 411– 25. https://doi.org/10.1080/17470210500151436

Nozick, R (1981). *Philosophical explanations*. Harvard University Press.

Olsson, E. (2015). Gettier and the method of explication: A 60-year-old solution to a 50-year-old problem. *Philosophical Studies*, *172(*1), 57-72. https://doi.org/10.1007/s11098-014-0383-z

Parfit, D. (1997). Reasons & motivation. *Proceeding of the Aristoteleian Society*, *Supplementary Volumes, 71*: 99-130. 10.1111/1467-8349.00021

Partee, B. (1986). Ambiguous pseudoclefts with unambiguous be. *Proceedings of NELS 16*, 354–66.

Penn, D.C., Holyoak K.J. & Povinelli, D.J. (2008). Darwin's mistake: explaining discontinuity between human and non-human minds. *Behavioral and Brain Sciences*, *31*, 109–78. https://doi.org/10.1017/S0140525X08003543

Pennington, N. & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition,* 49(1 – 2),123. https://doi.org/10.1016/0010-0277(93)90038-W

Pollock, J. (1979). A plethora of epistemological theories. in G. Pappas (ed.) *Knowledge and Justification*(93–115).. Dordrecht.

Pollock, J. (1986). *Contemporary theories of knowledge* (*1st Edition*). Rowman & Littlefield.

Pollock, J. & Cruz, J. R. (1999). *Contemporary theories of knowledge (2nd Edition)*. Rowman & Littlefield.

Premack, D. & Woodruff, A.J. (1978). 'Does the Chimpanzee Have a Theory of Mind' Revisited. In R. Byrne. and A. Whiten (eds.) *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans* (pp.160-79). Oxford University Press.

Pritchard, D. (2005). *Epistemic Luck.* Oxford University Press.

Pritchard, D. (2007). Anti-luck Epistemology. *Synthese*, *158*(3), 277–97. https://doi.org/10.1007/978-1-137-52692-2_2

Pritchard, D. (2008). McDowellian neo-mooreanism. A. Haddock. & F. MacPherson (eds.) *Disjunctivism: Perception, action, knowledge* (pp.283-310). Oxford University Press.

Pritchard, D. (2012). Anti-Luck Virtue Epistemology. *The Journal of Philosophy,109*(3), 247–79.

Pritchard, D. (2018). Epistemological disjunctivism & the biscopic treatment of scepticism. In V. Mitova. (ed.) *The factive turn in Epistemology* (pp. 15-32). Cambridge University Press.

Pryor, J. (2000). The skeptic and the dogmatist. Noûs, *34*(4), 517–549. DOI: 10.1111/0029-4624.00277

Pryor, J. (2007). Reasons and that-clauses. *Philosophical Issues*, *17(*1), 217-244. DOI: 10.1111/j.1533-6077.2007.00131.x

Putnam, H. (1975). The meaning of 'meaning'. *Minnesota Studies in the Philosophy of Science. 7,* 131-193.

Quine, W. (1951). Two dogmas of empiricism". *Philosophical Review*, *60*(1), 20–43.

Ratner, H. & Hill, L. (1991). Regulation and representation in the development of children's memory. Paper Presented to the Society for Research in Child Development, Seattle.

Reed, B. (2009). A new argument for skepticism. *Philosophical Studies*, *142*(1), 91-104. DOI: 10.1007/s11098-008-9299-9

Reed, B. (2013). Fallibilism, epistemic possibility and epistemic agency. *Philosophical Issues*, *23*, 40-69. DOI: 10.1111/phis.12003

Reed, B. (2016). Who Knows? in M.A. Fernandez (ed.) *Performance Epistemology* (pp.106-23). Oxford University Press.

Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H. & Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction*, 11(3–4), 347–64. DOI: https://doi.org/10.1207/s1532690xci1103&4_11

Rips, L. J. (2002). Circular reasoning. *Cognitive Science*, *26*(6), 767–95. https://doi.org/10.1207/s15516709cog2606_3

Rizzieri, A. (2011). Evidence does not equal knowledge. *Philosophical Studies*, *153(*2), 235–42. https://doi.org/10.1007/s11098-009-9488-1

Romero, M. (2005). Concealed questions and specificational subjects. *Linguistics & Philosophy 28*, 687-737. DOI: 10.1007/s10988-005-2654-9

Rorty, R. (1979). *Philosophy & the mirror of nature*. Princeton University Press.

Rosenberg, J. (2002). *Thinking about knowing*. Oxford University Press.

Ross. W.D. (1930). *The right and the good.* Oxford University Press.

Rysiew, P. (2011). Making it Evident: Evidence and evidentness, justification and belief. In T. Dougherty (ed.) *Evidentialism & its discontents* (pp. 207-25). Oxford: Oxford University Press.

Rysiew, P. (2018). Facticity and evidence. In V. Mitova. (ed.) *The factive turn in epistemology* (pp. 50-65). Oxford: Oxford University Press.

Russell, B. (1912) *The problems of philosophy*. Oxford University Press.

Saunders, J. Turk & Champawat, N. (1964). Mr. Clark's definition of 'knowledge'. *Analysis*, *25*(1), 8-9.

Scanlon, T.M. (1998) *What we owe to each other*. Belknap Press.

Schelling, T.C. (1960). *The strategy of conflict*. Harvard University Press.

Schnee, I. (2014). There is no knowledge from falsehood. *Episteme*, *12*(1), 53–74. DOI: 10.1017/epi.2014.26

Schiffer, S. (2009). Evidence = knowledge: Williamson's solution to skepticism. In P. Greenough & D. Pritchard (eds.). *Williamson on knowledge* (pp. 183-203). Oxford University Press.

Schlenker, P. (2003). Clausal equations (a note on the connectivity problem. *Natural Language and Linguistic Theory, 21*, 157–214. https://doi.org/10.1023/A:1021843427276

Scott-Phillips, T. (2015). *Speaking our minds*. Palgrave-Macmillan.

Schroeder, M. (2007). *Slaves of the passions*. Oxford University Press.

Schroeder, M. (2008). Having reasons. *Philosophical Studies*, *139*(1), 57–71. DOI: 10.1007/s11098-007-9102-3

Schroeder, M. (2015). Knowledge is belief for sufficient (objective and subjective) reason. In T. S. Gendler and J. Hawthorne (eds.) *Oxford Studies in Epistemology, volume 5***.** (pp. 226-252). DOI: 10.1093/acprof:oso/9780198722762.003.0008

Schroeder, M. (2021). *Reasons first*. Oxford University Press.

Schelling, T.C. (1960). *The strategy of conflict*. Harvard University Press.

Searle, J. (1969). *Speech acts: An essay in philosophy of language*. Cambridge University Press.

Searle, J. (1995). *The construction of social reality.* New York: Free Press.

Searle, J. (2001). *Rationality in action*. M.I.T. Press.

Sellars, W. (1956/1997). *Empiricism and the philosophy of Mind*. Harvard University Press.

Shope, R. (1983) *The analysis of knowing: A decade of research.* Princeton University Press.

Silins, N. (2005). Deception and evidence. *Philosophical Perspectives*, *19*(1), 375–404. 10.1111/j.1520-8583.2005.00066.x

Simion, M. (2018). Knowledge first functionalism. *Philosophical Issues*, *29(*1), 254–67. DOI: 10.1111/phis.12152

Simion, M. (2021). Knowledge and reasoning. *Synthese*, *199*(3), 10371-10388. DOI: 10.1007/s11229-021-03251-z

Simion, M. (2023). Resistance to evidence and the duty to believe. *Philosophy & Phenomenological Research*, *108*(1). DOI: 10.1111/phpr.12964

Smith, J.D., Couchman, J.J., & Beran, M.J. (2014). Animal metacognition: A tale of two comparative psychologies. *Journal of Comparative Psychology, 128*(2), 115–31. DOI: 10.1037/a0033105

Smithies, D. (2012). Mentalism and epistemic transparency. *Australasian Journal of Philosophy, 90*(4), 723–741. DOI: 10.1080/00048402.2011.627925

Smithies, D. (2012). Moore's paradox and the accessibility of justification. *Philosophy and Phenomenological Research, 85*(2), 273–300.

Smithies, D. (2014). The phenomenal basis of epistemic justification. In J. Kallestrup & M. Sprevak, (eds.) *New Waves in Philosophy of Mind* (pp. 98-124). Palgrave MacMillan.

Smithies, D. (2015). Why Justification Matters. In D. Henderson & J. Greco (eds.) *Epistemic Evaluation: Purposeful Epistemology* (pp. 224-44). Oxford University Press.

Smithies, D. (2016). Reflection on: On reflection. *Analysis, 76(*1), 55–69. DOI: 10.1093/analys/anv032

Snedegar, J. (2021). Competing reasons. In J. Brown & M. Simion (eds.) *Reasons, justification and defeat* (pp. 247-69). Oxford University Press.

Sosa, E. (1991). *Knowledge in perspective: Selected essays in epistemology*. Cambridge University Press.

Sosa, E. (2003). In L. BonJour & E. Sosa (authors) *Epistemic justification*: *Internalism vs. externalism, foundations vs. virtues.* Oxford University Press.

Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge, volume 1*. Clarendon Press.

Stanley, J. (2005). *Knowledge and practical interests.* Oxford University Press.

Sterelny, K. (2012). *The evolved apprentice*. Oxford University Press.

Steup, M. (1999). *An introduction to contemporary epistemology*. Pearson.

Sutton, J. (2007). *Without justification*. MIT Press.

Swain, M. (1974). Epistemic defeasibility. *American Philosophical Quarterly*, *11*, 15–25.

Swinburne, R. (2001) *Epistemic justification*. Oxford University Press.

Swinburne, R. (2011). Evidence. In T. Dougherty (ed.) *Evidentialism & its discontents* (pp. 194-206). Oxford University Press.

Thompson, V. A., Evans, J. St. B. T. & Handley, S. J. (2005). Persuading and dissuading by conditional argument. *Journal of Memory and Language, 53*(2), 238– 7.

Thune, M. (2010). Partial defeaters and the epistemology of disagreement. *The Philosophical Quarterly, 60*(239), 355–72. DOI: 10.1111/j.1467-9213.2009.611.x

Tinbergen, N. (1951). *The study of instinct*. Oxford University Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Harvard University Press.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. (2008). *Origins of human communication*. MIT Press.

Tomasello, M. (2014). *A natural history of human thinking.* Harvard University Press.

Tomasello, M. (2021). *Becoming human: A theory of ontogeny*. Harvard University Press.

Tomasello, M. & Call, J. (1997). *Primate cognition*. Oxford University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–91. DOI: 10.1017/S0140525X05000129

Turri, J. (2009). The ontology of epistemic reasons. *Noûs, 43*(3), 490–512. DOI: 10.1111/j.1468-0068.2009.00715.x

Turri, J. (2010). On the relationship between propositional and doxastic justification. 80 (2), 312-26. DOI: 10.1111/j.1933-1592.2010.00331.x

Turri, J. (2011). Manifest failure: The gettier problem solved. *Philosophers' Imprint*, *11*(8), 1–11.

Turri, J. (2012). In Gettier's Wake. In S. Hetherington (ed), *Epistemology: The key thinkers* (pp. 214-29). Continuum.

Turri, J. (2016). Knowledge as achievement, more or less. in M. Vargas (ed) *Performance Epistemology* (pp. 124-34). Oxford University Press.

Unger, P. (1968). An analysis of factual knowledge. *Journal of Philosophy*, *65*(6), 157–70.

Warfield, T. (2005). Knowledge from falsehood. *Philosophical Perspectives, 19*(1), 405–16. DOI: 10.1111/j.1520-8583.2005.00067.x

Washburn, D. A., Smith, J. D., & Shields, W. E. (2006). Rhesus monkeys (macaca mulatta) immediately generalize the uncertain response. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(2), 185–189. https://doi.org/10.1037/0097-7403.32.2.185

Weatherson, B. (2003). What good are counterexamples? *Philosophical Studies*, *115*(1), 1–31. DOI: 10.1023/a:1024961917413

Wedgwood, R (2002). Internalism explained. *Philosophy and Phenomenological Research*, *65*(2), 349–369. DOI. 10.1111/j.1933-1592.2002.tb00206.x

Weinberg, J. (2017). Knowledge, noise and curve-fitting: A methodological argument for justified true belief? In R. Borges, C. de Almeida, P. Klein (eds.) *Explaining knowledge: New essays on the gettier problem* (pp. 253–72). Oxford University Press.

Weinstock, M., Neuman, Y. & Tabak, I. (2004). Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology,* 29(1), 77–94. https://doi.org/10.1016/S0361-476X(03)00024-9

Whiten, A. & Byrne, R. W., (eds.) (1997). *Machiavellian intelligence II: Extensions and evaluations*. Cambridge University Press.

Williams, B. (1981). *Moral luck*. Cambridge University Press.

Williams, E. (1983). Semantic versus syntactic categories. *Linguistics & Philosophy,* 6, 426–46. https://doi.org/10.1007/BF00627484

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Wright, C. (1991). Scepticism and dreaming: Imploding the dream. *Mind*, *397*, 87–115. https://doi.org/10.1093/mind/C.397.87

Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–65. https://doi.org/10.1038/nn.4244

Yong Lee, K. (2021). Reconsidering alleged cases of knowledge from falsehood. *Philosophical Investigations, 44*(2), 151–62.

Zagzebski, L. (1994). The inescapability of gettier problems. *Philosophical Quarterly*, *44*(174), 65–73.

Zagzebski, L. (2017). The lesson of gettier. in R. Borges, C. de Almeida, & P. Klein (eds.) *Explaining knowledge: New essays on the gettier problem* (pp. 179–91). Oxford University Press.

Zhao, B. (2022). Knowledge from falsehood, ignorance of necessary truths and safety. *Philosophia*, *50*, 833–45. https://doi.org/10.1007/s11406-021-00410-x