



Mortini, Dario (2022) *Knowledge and legal proof between modality and explanation*. PhD thesis.

<https://theses.gla.ac.uk/83179/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study,  
without prior permission or charge

This work cannot be reproduced or quoted extensively from without first  
obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any  
format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author,  
title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

# **Knowledge and Legal Proof Between Modality and Explanation**

**Dario Mortini**

A thesis presented for the degree of  
Doctor of Philosophy



School of Humanities

College of Arts

University of Glasgow

8th of July 2022

# Dissertation Outline

I begin my dissertation by charting and assessing two competing approaches to theorise about the nature of knowledge – modalism and explanationism. According to the former, knowledge equates with a belief which is true in a relevant set of possible worlds; according to the latter, knowledge is a matter of believing the truth on the basis of the right explanation. When it comes to the competition between modalism and explanationism in traditional epistemology, I reject explanationism and I endorse modalism: I move novel objections against the most recent explanationist account of knowledge and I develop a novel version of modalism. This new version of modalism consists of a different safety condition which is informed by the etiological theory of functions made popular in the philosophy of biology. I argue that it does better than both extant versions of modalism and extant versions of explanationism.

I then move on to the contrast between modalism and explanationism in the context of applied legal epistemology. By emphasising a few structural connections between knowledge and legal proof, epistemologists have moved quite freely from a modalist account of knowledge to a modalist account of legal proof. A case in point is Duncan Pritchard, who proposes a modal condition for knowledge (safety) and then extends it to legal standards of proof. On this second issue, I see things differently: in legal epistemology, I endorse explanationism and I reject modalism. Accordingly, I object to influential modal accounts of legal proof and I offer an alternative explanationist account. I develop an account of legal proof as an inference to the best contrastive explanation: by drawing on up-to-date literature on abductive reasoning in the philosophy of science, I bring into focus hitherto unappreciated connections between the nature of legal proof and the structure of scientific explanation. The final upshot is a well-informed account of legal proof which does away with modal conditions and better captures the nature of juridical proof.

My dissertation advances two closely related debates by emphasising a neat separation between modalism and explanationism: while the former is the right approach to theorise about the nature of knowledge, only the latter can provide a satisfactory account of legal proof. Such separation offers a new picture of the relationship between knowledge and legal proof, and it provides new insights into several debates at the intersection of traditional and applied legal epistemology.

# Contents

<b>1</b>	<b>Setting the Stage</b>	<b>11</b>
1.1	Overview of the Chapter . . . . .	11
1.2	Gettierology and Metaepistemology . . . . .	11
1.3	Modalism: Motivations and Taxonomy . . . . .	15
1.4	Explanationism: Motivations and Taxonomy . . . . .	18
1.5	The Plan . . . . .	19
1.6	Chapter Summary . . . . .	20
<b>2</b>	<b>A Closer Look</b>	<b>21</b>
2.1	Overview of the Chapter . . . . .	21
2.2	Compromising: Virtue-reliabilist Aptness . . . . .	21
2.3	Between Modalism and Explanationism . . . . .	23
2.4	Fine-grained Versions of Safety . . . . .	24
2.4.1	Standard Method-relative Safety . . . . .	25
2.4.2	Ability-relative SSS Safety . . . . .	27
2.5	Problems with Safety (and Explanationist Solutions) . . . . .	29
2.6	The Way Forward . . . . .	31
2.7	Chapter Summary . . . . .	32
<b>3</b>	<b>Solving the Safety Dilemma</b>	<b>33</b>
3.1	Overview of the Chapter . . . . .	33
3.2	The Hardest Challenge for Safety . . . . .	33
3.3	Belief-forming Methods . . . . .	36
3.4	Environments . . . . .	39
3.5	Environment-relative Safety . . . . .	42
3.6	Safety as a Conjunctive Condition . . . . .	43
3.7	A New Solution to the Safety Dilemma . . . . .	44
3.8	Theoretical Virtues . . . . .	47
3.9	Chapter Summary . . . . .	48
<b>4</b>	<b>Against Kairetic Explanationism</b>	<b>49</b>
4.1	Overview of the Chapter . . . . .	49
4.2	The Competition: Safety and Kairetic Explanationism . . . . .	49

4.3	Kairetic Explanationism Outlined . . . . .	51
4.4	Kairetic Explanationism Exposed: Stopped Clocks . . . . .	53
4.5	Kairetic Explanationism Exposed: Fake Barns . . . . .	55
4.6	Environment-relative Safety Defended . . . . .	59
4.7	Chapter Summary . . . . .	63
<b>5</b>	<b>Modal Proper Functionalism Defended</b>	<b>64</b>
5.1	Chapter Overview . . . . .	64
5.2	A New Hybrid . . . . .	64
5.3	The Rise and Fall of Safety . . . . .	65
5.4	Diagnosis and Desiderata . . . . .	68
5.5	Modal Proper Functionalism Outlined . . . . .	70
5.6	Modal Proper Functionalism in Action . . . . .	74
5.6.1	Gettier Cases: Clocks and Barns . . . . .	74
5.6.2	Necessity . . . . .	76
5.6.3	Sufficiency . . . . .	78
5.7	Chapter Summary . . . . .	80
<b>6</b>	<b>Modalism vs Explanationism: Legal Proof</b>	<b>82</b>
6.1	Overview of the Chapter . . . . .	82
6.2	Applied Legal Epistemology . . . . .	82
6.3	Meta-epistemological Desiderata for Legal Epistemology . . . . .	84
6.4	Legal Proof and Proof-Paradoxes . . . . .	85
6.5	Explanatory, Descriptive and Normative Projects . . . . .	89
6.5.1	Case Study I: Individualised versus Statistical Evidence . . . . .	90
6.5.2	Case Study II: The Epistemic Normativity of Criminal Conviction . . . . .	92
6.6	Swapping: Against Modalism and For Explanationism . . . . .	93
6.7	Chapter Summary . . . . .	94
<b>7</b>	<b>Against Modal Accounts of Legal Proof</b>	<b>95</b>
7.1	Overview of the Chapter . . . . .	95
7.2	Individualised Evidence and Luck . . . . .	95
7.3	The Puzzle of Statistical Evidence . . . . .	97
7.4	A Common Anti-luck Condition? . . . . .	100
7.5	Causation . . . . .	102
7.6	Modal Conditions . . . . .	105
7.6.1	Sensitivity . . . . .	105
7.6.2	Safety . . . . .	108
7.7	Diagnosis . . . . .	112
7.8	Chapter Summary . . . . .	115

<b>8</b>	<b>Legal Proof as Inference to the Best Contrastive Explanation</b>	<b>116</b>
8.1	Overview of the Chapter . . . . .	116
8.2	Pritchard’s Safety Account of Legal Proof . . . . .	116
8.3	Two Counterexamples, One Diagnosis . . . . .	119
8.4	Legal Proof as Inference to the Best Contrastive-why Explanation . . . . .	120
8.5	The Linguistic Argument . . . . .	123
8.6	The Psychological Argument . . . . .	124
8.7	Coda: Norms of Criminal Conviction and Explanation . . . . .	126
8.8	Chapter Summary . . . . .	127
<b>9</b>	<b>Conclusions</b>	<b>128</b>
9.1	Knowledge and Safety . . . . .	128
9.2	Knowledge and Legal Proof . . . . .	129
<b>A</b>	<b>Swampman Cases</b>	<b>130</b>
<b>B</b>	<b>Epistemic Supervenience and the Analysis of Knowledge</b>	<b>134</b>
	<b>Bibliography</b>	<b>150</b>

# Dedication

*To Sabina, whom I love (always and forever)*

# Acknowledgements

This thesis would not have been possible without the support of many people.

I am very grateful to my first supervisor, Chris Kelp. In the last four years, he has consistently shown support and encouragement. Both professionally and personally, he has set an invaluable example for me. My thanks to him for his unending patience, kindness and help.

I am also very grateful to my second supervisor, Adam Carter. Like Chris, he has encouraged me and helped me at every stage of my PhD, and taught me a lot (perhaps more than I could possibly thank him for). Adam has always answered to my (mostly silly) emails in a quick and prompt manner. He made me feel cared for and supported, and for this I am very grateful.

I also want to thank other staff members at Glasgow for making my time there worthwhile and rewarding. Thanks to Mona Simion for giving me hope and optimism when it comes to the philosophy job market, and for encouraging me to read the literature on etiological proper function. Thanks to Jack Lyons for answering to all my questions about process-reliabilism. Thanks to Stephan Leuenberger for encouraging me to talk and write about (some) metaphysics, and for urging me to cultivate my interests in legal philosophy. Thanks to David Bain and Ben Colburn, two caring and kind heads of department. Thanks to Neil McDonnell for reassuring me that using “outwith” in a sentence is acceptable. Thanks to Gary Kemp and Alan Weir for making me feel welcome, and for teaching me how to talk philosophy while sitting in a pub. Thanks to Adam Rieger for showing me that it’s actually possible to teach logic in Microsoft Word. Thanks to George Pavlakos for drinking coffee with me in a dreich December day and for taking the time to talk to me about legal philosophy. Thanks to Chris Wyllard-Kyle for being someone to look up to: brilliant, available and generous. Thanks to Jennifer Corns, for teaching me how to do the hardest thing (how to teach philosophy), and for infusing with happiness and enthusiasm every conversation I’ve had with her.

I’m also grateful to the people I’ve met in Cologne during my visiting. Thanks to Sven Bernecker for providing detailed comments to my drafts and for showing me how to move sharp objections with kindness. Thanks to Paul Silva, Wes Siscoe and Luis Rosa for making me feel welcome at CONCEPT and for giving me a fresh perspective on many philosophical issues we discussed together. Thanks to Thomas Grundmann for his generous feedback on my Brown Bag talk. Thanks to Anna Maria Asunta Eder for teaching me a lot by asking the right questions during the philosophy seminars.

Thanks also to Christine LeRiguer, Patricia McLaren, Clare Smith, Leslie Watson and Catherine Hastie for all their help with the bureaucracy side of things, especially during COVID times. I could not have made it without their timely and professional support.

Thanks to the undergraduate students that attended my seminars and guest lectures: I'm especially grateful to Charlotte Connell, Jasmine Hunt, Marlies Humpelstetter, Bonnie Campbell, Tom Flohrer, Andrew Eugenicos, Monica Tate, Nicola McLaughlin, Alice Caunt and Jim Mumba. Their questions pushed me to clarify the material as much as possible, and discussing philosophy with them was a real privilege and delight.

Heartfelt thanks to the grad students I've crossed paths with in Glasgow and elsewhere. Thanks to Joaquim Giannotti for being a friend and a mentor. Thanks to Laura Fearnley for lighten many rainy days in the department office with her sense of humour. Thanks to Ross Patrizio for being the coolest epistemology grad student ever, and for talking with me about two topics I love (indie rock and epistemology). Thanks to Lilith Newton for her contagious enthusiasm, and for shaping my views in legal epistemology with her sharp comments and criticisms. Thanks to Angie O' Sullivan for many helpful conversations and for showing me both the good and the bad sides of my ideas. Thanks to Giada Fratantonio for being one of my biggest sources of inspiration, both personally and philosophically. Thanks to Giorgia Foti for her instructive analogies between epistemology views and Italian politics. Thanks to Lewis Ross for sending me his drafts, for inspiring me to learn about evidence law and for sparking my interest in the proof paradox. Thanks to Marvin Backes for showing me the virtues of epistemic pluralism, for talking with me about closure and for telling me about the rivalry between Köln and Düsseldorf while we were cycling along the Rhine.

Thanks also to friends and family who have supported me along the way. Thanks to Luca Cereda, for being the best best friend one could possibly ask for. Thanks to Adriana Alcaraz-Sanchez: she's been not only a sharp and bright fellow grad student, but also the dearest, kindest and most trustworthy friend. Thanks to Nacho and Cristina, for being my two glowing suns in otherwise gloomy Scottish rainy days. Thanks to my sisters, Giulia and Emma, for their patience and understanding. Thanks to my mother for believing in me, always and no matter what.

Finally, special thanks to Sabina, for bringing joy into my life both within and outwith philosophy. This thesis is dedicated to her.

# **Author's Declaration**

I declare that all the work in this dissertation is the result of my own efforts, except where properly indicated by means of quotations and references. This work has not been submitted for any other degree of professional qualification. This work has received generous and invaluable financial support from the Scottish Graduate School for Arts and Humanities, the University of Glasgow College of Arts and the Royal Institute of Philosophy.

# Statement of Length

This dissertation, including footnotes and references, contains about 73 000 words. It meets the minimum word limit set by the University of Glasgow College of Arts (70 000) and does not exceed the maximum word limit (90 000) set by the University of Glasgow College of Arts.

# Published Material

At the time of the submission of this thesis (July 2022), portions of Chapters 3, 4 and 7 have been accepted for publication or published in the following journals:

- Mortini, Dario (2022). A new solution to the safety dilemma. *Synthese* 200 (2):1-17.
- Mortini, Dario (2022). The Explanationist and the Modalist. *Episteme*: 1-16.
- Mortini, Dario (forthcoming). Knowledge, Individualised Evidence and Luck. *Philosophical Studies*.

# Chapter 1

## Setting the Stage

### 1.1 Overview of the Chapter

This chapter introduces the methodological commitments of the dissertation, as well as the two main families of views that take centre stage in the subsequent chapters (modalism and explanationism). I clarify the subject matter of the epistemological questions that I aim to raise, and I also provide a taxonomy of different versions of modalism and explanationism. The chapter also serves as a self-standing overview of two competing approaches to theorising about the nature of knowledge (viz. modalism and explanationism).

### 1.2 Gettierology and Metaepistemology

Epistemology is commonly referred to as the theory of knowledge. In this chapter, I shall follow and adopt this broad characterisation but also add some suitable specifications. I am also interested in the nature of knowledge: more specifically, I am interested in which conditions – if any – can be necessary, sufficient and perhaps even necessary *and* sufficient for knowledge. I will assume without arguments that knowledge requires at least truth and belief as necessary conditions, and focus on the type of *connection* with truth that can be necessary and sufficient for knowledge.<sup>1</sup> This connection is purportedly captured by so called anti-Gettier conditions – the type of condition that rules out Gettier-style cases, such as the following:

**Sheep in the field.** Roddy is a farmer. One day he is looking into a field near-by and clearly sees something that looks just like a sheep. Consequently he forms a belief that there is a sheep in the field. Moreover, this belief is true in that there is a sheep in the field in question. However, what Roddy is looking at is not a sheep, but rather a big hairy dog that looks just like a sheep and which is obscuring from view the sheep standing just behind.<sup>2</sup>

---

<sup>1</sup>See Shope (1983) Ichikawa & Steup (2014) and Carter & Littlejohn (2021) for an overview. For the contention that knows doesn't require truth, see Hazlet (2010). For an objection against the necessity of belief for knowledge, see Radford (1966).

<sup>2</sup>Chisholm (1966).

**Fake Barn.** Using his reliable perceptual faculties, Barney forms a true belief that the object in front of him is barn. Barney is indeed looking at a barn. Unbeknownst to Barney, however, most objects that look like barns in these parts are in fact barn façades.<sup>3</sup>

In each of these cases originally proposed by Gettier (1963), the agent has a (justified) true belief, and yet they lack knowledge. But what kind of condition can predict absence of knowledge here? What kind of condition can capture the right kind of connection to truth that is necessary and sufficient for knowledge? In other words, which is the correct anti-Gettier condition? We may call this the ‘Gettier problem’, and subsume under the label ‘Gettierology’ all the attempts to answer to this question.<sup>4</sup> However, before engaging with this more specific epistemological question, it’s worth stepping back to clarify the substantial commitments and assumptions underlying it and do some important metaepistemology. Accordingly, let’s distinguish between the following two questions:

**Epistemological Question.** *Which condition* can capture the right connection to truth that is necessary and sufficient for knowledge?

**Metaepistemological Question.** *How* can such condition be individuated?

As some commentators have noticed, epistemologists often remain silent on these core metaepistemological questions.<sup>5</sup> The first question constitutes an attempt to provide an informative, non-circular *analysis* of knowledge of the form “S knows that *p* if and only if: ”. The second question asks what exactly is being analysed and how to best do so. Let’s focus on the metaepistemology question. Following Sosa (1991, 2015, 2021), we can distinguish between three types of analyses depending on the nature of the *analysandum*:

- A *semantic analysis* of the word knows.
- A *conceptual analysis* of the concept knows.
- A *metaphysical analysis* of the objective phenomenon of knowing.

Semantic analyses share the methodological assumption that knowledge attributions are a useful guide to the nature of knowledge (e.g., Stanley 2005). Conceptual analyses purport to investigate the nature of knowledge by looking at the features of our concept of

---

<sup>3</sup>Goldman (1976).

<sup>4</sup>It is unclear whether the Gettier problem is a new problem. As Plantinga (1993) writes: ‘Of course there is an interesting historical irony here: it isn’t easy to find many really explicit statements of a JTB analysis of knowledge prior to Gettier. It is almost as if a distinguished critic created a tradition in the very act of destroying it.’ On this point, see also Antognazza (2015 and Dutant (2015).

<sup>5</sup>For example, Shope (1983: 34) writes that “lack of clarity in talking about concepts, and widespread contemporary suspicion of the analytic/synthetic distinction have led most authors to avoid committing themselves to the view that what is being analyzed is the content of the concept *knowledge*”. In a similar vein, Ichikawa and Steup (2018) note that “In practice, many epistemologists engaging in the project of analyzing knowledge leave these metaphilosophical interpretive questions unresolved; attempted analyses, and counterexamples thereto, are often proposed without its being made explicit whether the claims are intended as metaphysical or conceptual ones.”

knowledge (e.g., Jackson 1998) and by providing a type of conceptual analysis (a common methodology in philosophy). Differently, metaphysical analyses are concerned with the referent of our word and concept of knowledge. As Sosa (2015: Chapter 1) is keen to emphasise:

Here our focus is on an *objective phenomenon* that need be neither expression nor concept. Our focus is rather on a state that people host or an act that they perform. This is the phenomenon whose ontology we now wish to understand. What is the nature of such a state or act, and how is it grounded? In virtue of what is it actual when it is actual?<sup>6</sup>

Another author who's explicitly concerned with the referent and the *phenomenon* of knowledge is Ruth Millikan:

(...) I propose to describe knowledge in the same sort of spirit in which science has described, say, tuberculosis or acidity or genes. Whatever else it is, knowledge surely is a phenomenon in the world that we need to understand. (Millikan 1993: 243)

Back to our metaepistemological question: in this and in the subsequent chapters, I will be concerned first and foremost with a metaphysical analysis of knowledge and focus on the objective worldly phenomenon of knowing. As a result, I will have to set aside (interesting) linguistic questions concerning knowledge attributions. However, I will not completely discard the weight of broadly conceptual considerations: I take our concept of knowledge to have important consequences for our investigation of it. In sum, I will embrace a somewhat *pluralistic metaepistemology*: while my main focus will be on its referent, I shall pay attention also to some features of the concept of knowledge (for instance, when it comes to motivate the incompatibility between knowledge and luck). This pluralistic metaepistemology will constrain the philosophical methodology of the subsequent chapters. To highlight important structural features of knowledge, I will construct and appeal to hypothetical cases: these cases elicit specific intuitions that I take to be a reliable – though certainly not infallible – guide to investigate the nature of knowledge. To get a better sense of this pluralistic metaepistemology, it will be useful to compare and contrast it with further metaepistemological views. Jonathan Ichikawa and Carrie Jenkins also advocate a type of pluralism, and they're interested in both the concept and the referent of knowledge. As they nicely put it:

We do not assume that it is metaphysical, rather than representational, questions that are central to epistemology. There is a clear tradition of epistemological investigation of the nature of the concept KNOWS, the semantics of 'knows', and the psychology of knowledge ascriptions. We are pluralists about epistemology; we see no obvious reason not to think these are of significant epistemological interest in their own right, alongside questions about knowledge itself. (Ichikawa & Jenkins 2017: 127)

---

<sup>6</sup>Italics are mine. Also quoted in Carter & Sosa (2022).

The metaepistemological approach endorsed here aligns well with Ichikawa and Jenkins' pluralism. Differently, it parts way with Kornblith's *naturalistic* metaepistemology. Kornblith (2002) holds that knowledge is a natural kind, and, as such, cannot be investigated relying on a priori methods or by investigating its concept. Because Kornblith's focus is *exclusively* on the referent rather than the concept of knowledge, it is in tension with the broadly pluralistic metaepistemology defended here. Ruth Millikan (1993: Chapter 12) also sets aside considerations pertaining to the concept or the word "knows". Here's a telling quote:

The setting within which I will explore the phenomenon of knowledge is that of evolutionary theory. Thus I make an immediate and obvious departure from conceptual analysis. Indeed, should the phenomena I discuss turn out to have no connection with the term "knowledge," that would surprise but not unduly pain me. They would still be phenomena deserving of philosophical reflection, interesting in their own right.

In dissertation, I part ways with this approach. Even if our main concern will be with the state of knowledge, conceptual considerations will have at least *some* weight.

So much for our epistemology and metaepistemology: we are after a metaphysical analysis of knowledge in the form of an anti-Gettier condition that can be necessary and sufficient for knowledge. However, the idea that knowledge can be (metaphysically) analysed faces a powerful objection from Williamson's knowledge-first epistemology (Williamson 2000). The core metaepistemological idea of knowledge-first epistemology is to take knowledge as basic and use it to elucidate further philosophically relevant phenomena (e.g., belief, justification, evidence, understanding, assertion). Williamson provides different arguments in support of his metaepistemological position: a pessimistic abduction based on the failure of providing a satisfactory anti-Gettier condition, and an argument against the very possibility of equating the concept of knowledge with a distinct concept.<sup>7</sup> While intrinsically compelling and potentially far-reaching, this metaepistemological position can be resisted. As McGlynn notices (2014), it may be too early to throw in the towel when it comes to provide a satisfactory (metaphysical) analysis of knowledge. In his own words:

I don't think we should be impressed by the inductive argument against the possibility of an analysis of knowledge. Stating necessary or sufficient conditions for anything philosophically interesting is hard, and it remains so even once one gives up the reductive ambitions that many of the participants in the post-Gettier debate seem to have harboured. That there's a distinctive problem faced by proponents of the thesis that knowledge is true belief that meets further conditions hasn't been demonstrated. (McGlynn 2014: 174)

Indeed, the two approaches under consideration here (modalism and explanationism) are promising candidates to spell out this further anti-Gettier condition. In the remainder

---

<sup>7</sup>See the exchange between Cassam (2009) and Williamson (2009a) for relevant discussion of these arguments.

of the chapter I shall provide a rough characterisation of each, and taxonomise different versions of both modalism and explanationism.<sup>8</sup>

### 1.3 Modalism: Motivations and Taxonomy

Modalist anti-Gettier conditions begin with a specific diagnosis of the kind of accidentality standardly found in Gettier-style cases. Their key insight is that in these cases the agent's true belief is only *luckily* true. Back to the cases mentioned above. Roddy's belief is true only by luck: only luckily there happens to be a sheep behind the rock that makes the belief under consideration true. The same holds for Barney: only luckily he picks the real barn amidst a forest of fakes. This suggests the following:

**Anti-luck platitute.** Knowledge and luck are essentially incompatible.<sup>9</sup>

It is important to clarify the specific type of luck which is incompatible with knowledge. You may be lucky to have a proper-functioning visual system when you gaze out the window and see a blue sky, but this doesn't prevent you from coming to know that the sky is blue. Similarly, Nancy Drew may be lucky in finding a key piece of evidence against the murderer of the cases she's working on, but this doesn't prevent her from coming to know who committed the murder. Following Pritchard (2005, 2007, 2012a 2013), the type of luck incompatible with knowledge is best called *veritic* luck – the type of luck which makes a belief only accidentally *true*.

With the anti-luck platitute on the table, advocates of modalism can then advance and develop a suitable *anti-luck condition* which is proposed as necessary or sufficient for knowledge depending on the modalist anti-Gettier condition in question. To constitute knowledge, a true belief must meet such anti-luck condition which is meant to rule out Gettier-style cases. Crucially for our purposes, such condition is standardly cast in terms of *possible worlds*. Accordingly:

**Modalist core commitment.** Knowledge is a *modal* matter: whether the true belief under consideration counts as knowledge depends on whether the true belief under consideration is true in a relevant set of possible worlds.

There's a number of different ways to spell out the modalist core commitment: modalist have an internal disagreement on which possible worlds are relevant for knowledge, or on how to best formulate their modal anti-Gettier condition. In what follows, I provide a taxonomy of the most influential modalist anti-Gettier conditions.

---

<sup>8</sup>One more word of caution on the knowledge-first approach to epistemology. In Appendix B, I consider and reject a metaphysical objection against the viability of a knowledge-first approach to epistemology. But even if this objection fails to go through, adopting the knowledge-first approach involves a number of methodological commitments that I shall not take on in this thesis.

<sup>9</sup>See Unger (1968) and Pritchard (2005, 2007, 2012a 2013) for a canonical defense of this platitute. See Carter & Berrocal (2017) for a thorough discussion of the anti-luck platitute. See Hetherington (2013) for the contention that knowledge and luck are actually compatible.

One way to develop a modalist anti-Gettier condition is in terms of *subjunctive conditionals*, i.e. sentences of the form ‘if ... were (not) the case, then... would (not) be the case’. In symbols:  $A \rightarrow B$ , where  $A$  and  $B$  represent the antecedent and the consequent respectively. For example: ‘if the glass was to fall, it would break’. Standard truth-conditions for subjunctive conditionals are cast in terms of possible worlds:

$A \rightarrow B$  is true at a world  $w$  if and only if all the worlds that are most similar to  $w$  where  $A$  is true are worlds where  $B$  is true as well. (Comesaña 2007: 782)

Subjunctive conditionals can yield powerful modalist anti-Gettier conditions. These are also called *tracking* conditions: a true belief tracks the truth and can therefore exclude knowledge-undermining luck. Nozick (1981) proposes the following modalist anti-Gettier condition:

S knows that  $p$  if and only if:

- (1) S believes  $p$ .
- (2)  $p$  is true.
- (3) had  $p$  been false, then S would not have believed  $p$  (variation).
- (4) had  $p$  been true, then S would have believed  $p$  (adherence).

Taken together, clauses (3) and (4) constitute the *sensitivity* condition. Interestingly, such condition can explain absence of knowledge in the Gettier-style cases outlined above: in fact, the variation clause is violated. Had there not been a sheep behind the rock, Roddy would have believed it anyway. Had there not been a real barn (suppose that there was a fake instead), Barney would have believed it anyway. Because the third condition is not met, knowledge is absent.

The sensitivity condition is, however, much too strong. Firstly, it doesn’t permit to know the denial of sceptical hypotheses: if you were a radically deceived subject (e.g., a brain in a vat), you’d believe it anyway. This leads to problematic closure failures and abominable conjunctions (I know that I have a hand, but I don’t know that I’m not a handless brain in a vat). Take the following suitably restricted closure principle on knowledge:

**Closure.** If one knows  $p$  and competently deduces  $q$  from  $p$ , thereby coming to believe  $q$  while retaining one’s knowledge that  $p$ , then one comes to know that  $q$ . (Williamson 2000; Hawthorne 2005)

Let  $p$  and  $q$  be the following propositions: that you have a hand and that you are not a brain in a vat. While you can competently deduce  $q$  from  $p$ , only the former is sensitive. This yields the kind of abominable conjunction described by DeRose (1995).<sup>10</sup> Moreover, sensitivity rules out common cases of inductive knowledge. Sosa (1999) proposes the following case of knowledge which nonetheless violates the sensitivity condition:

---

<sup>10</sup>Bernecker (2012) argues that closure considerations can’t be used to argue in favour of safety and sensitivity. Roush (2005) develops a version of sensitivity that complies with principle like deductive closure.

**Trash Chute.** On my way to the elevator I release a trash bag down the chute from my high rise condo. Presumably I know my bag will soon be in the basement. But what if, having been released, it still (incredibly) were not to arrive there? That presumably would be because it had been snagged somehow in the chute on the way down (an incredibly rare occurrence), or some such happenstance. But none such could affect my predictive belief as I release it, so I would still predict that the bag would soon arrive in the basement. My belief seems not to be sensitive, therefore, but constitutes knowledge anyhow, and can correctly be said to do so. (Sosa 1999: 146-147)

To do better, Sosa proposes to modify Nozick's tracking theory by appealing to a different subjunctive conditional:

S knows that  $p$  if and only if:

- (1) S believes  $p$ .
- (2)  $p$  is true.
- (3) Were S to believe  $p$ , then  $p$  would be true.

This is a different subjunctive conditional, that Sosa called 'Safety'. Because subjunctive conditionals do not contrapose, sensitivity and safety are distinct.<sup>11</sup> In fact, safety permits to know the denial of sceptical hypotheses, complies with plausible closure principles and also accommodates cases like **Trash Chute**.

The safety condition became the most popular modalist condition. What is more, the safety condition has been further developed without explicit reference to subjunctive conditionals.<sup>12</sup> For example, Williamson (2000) and Pritchard (2007) propose the following formulations of safety:

**Williamson's Safety.** If one knows, one could not easily have been wrong in a similar case (Williamson 2000: 147)

**Pritchard's Safety.** S's belief is safe if and only if in most nearby possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, and in all very close nearby possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, the belief continues to be true. (Pritchard 2005, 2007)

Notice finally that safety also explains absence of knowledge in the Gettier-style cases mentioned above. Roddy could have easily formed the belief that there's a sheep in the

---

<sup>11</sup>Consider: If Jon had said he was going to the party, Sarah would have as well. But if Sarah hadn't said she was going to the party, Jon would have – he's been avoiding her all week.

<sup>12</sup>Another defender of safety is Sainsbury (1997). For a clear overview of the safety condition, see Rabinowitz (2011).

field by looking at the painted sheep: in close possible worlds, the belief under consideration is false. Ditto for Barney, who could have easily looked at a fake barn and formed a false belief. In close possible worlds, Barney's belief is also false.

Thus, to sum up, modalist approaches equate knowledge with a belief which is true in the relevant possible worlds. Among modalist approaches, we can distinguish:

- Subjunctive conditional tracking: sensitivity (Nozick 1981).
- Subjunctive conditional tracking: safety (Sosa 1999).
- Simple safety (Pritchard 2007, Williamson 2000).

These are the bare bones of modalism, but we will go back to safety and make more fine-grained distinction in Chapter 2. It is now time to turn the attention to the main competitor of modalism – explanationism.

## 1.4 Explanationism: Motivations and Taxonomy

Advocates of explanationism set out to capture the kind of accidentality at play in Gettier-style cases by focusing on (the lack of) explanatory connections between the belief in question and its truth. The focus on explanatory connections is especially important: the notion of explanation is in fact *hyperintensional* – it flank substitution *salva veritate* of necessarily equivalent and coextensive terms.<sup>13</sup> As such, explanation is a kind of notion that can't be adequately captured by modal notions. To better illustrate this point, consider the following pair of sentences:

- 1 Your friend is in danger *because* Dr. Jekyll is their neighbour.
- 2 Your friend is in danger *because* Mr. Hyde is their neighbour.

Dr. Jekyll and Mr. Hyde are necessarily equivalent and therefore coextensive terms: whatever is true of the former is also true of the latter in every possible world. Yet only [2] provides an explanation of why your friend is in danger. Modal notions struggle to accommodate this difference.<sup>14</sup>

Lack of appropriate explanatory connections help to shed light on what goes wrong in Gettier-style cases. What explains why the belief under consideration is hold is not its truth, but rather some *deviant* factor or some specific *interference*: Roddy believes that there's a sheep in the field because of a replica rather than because of the actual sheep. And perhaps Barney believes that there's a barn in front of him because the object merely looks like a barn just like other barn replicas look like barns. So far, so good.

Explanationists are free to develop their condition depending on their preferred reading of explanation – a notoriously opaque notion. Here's a (non-exhaustive) list of the most developed explanationist conditions:

---

<sup>13</sup>On the hyperintensionality of explanation, see Nolan (2014). On the hyperintensionality of explanation in the specific context of mathematics, see Baron et al. (2017 and Schnieder (2011).

<sup>14</sup>See Berto & Nolan (2021) for a thorough overview of hyperintensionality.

- 1 *Contextualist explanationism*. According to Neta (2002), the notion of explanation varies with context; this contextualist reading helps with Gettier-style cases and closure.
- 2 *Abductivist explanationism*. Rieber (1998) holds that the relevant type of explanation is an inference to the best explanation (abduction).
- 3 *Scopic explanationism*. Carrie Jenkins (2006) adopts a more sophisticated notion of explanation: she appeals to the truth of  $p$  being a good explanation for an external *observer* – someone not acquainted with a more specific set of facts.
- 4 *Kairetic explanationism*. Bogardus and Perrin (forthcoming) lean on Strevens kairetic test for scientific explanation (Strevens 2008). Complete explanations amounts to the set of *explanantes* that can't be subtracted away.
- 5 *Contrastivist explanationism*. According to Bernecker (forthcoming), the relevant notion of explanation is *contrastive*: S knows  $p$  if and only if the best explanation available favour  $p$  rather than  $q$ .

Summing up: explanationist approaches to the Gettier problem insist on explanatory connections between the belief under consideration and its truth. Different notions of explanation yield different explanationist conditions; however, each of these conditions rejects the key modalist assumption that knowledge is a matter of truth across possible worlds.

## 1.5 The Plan

Let's recap: so far, I've laid out two competing ways to capture the right anti-Gettier condition on knowledge – modalism and explanationism. The former focusses on relevant possible worlds; the latter on explanatory connections between the belief under consideration and its truth. With these points in play, I now give a more detailed overview of the main thesis defended in this dissertation, and also a more precise outline of the upcoming chapters.

When it comes to the competition between modalism and explanationism in traditional epistemology, I reject explanationism and I endorse modalism. In chapter 3, I rescue the safety condition from a thorny dilemma that casts doubt on its necessity for knowledge (and its prospects as anti-Gettier condition). In Chapter 4, I move novel objections against the most recent explanationist account of knowledge (kairetic explanationism) and next, in Chapter 5, I develop a novel version of modalism. This new version of modalism consists of a different safety condition which is informed by the etiological theory of functions made popular in the philosophy of biology. I argue that it does better than both extant versions of modalism and extant versions of explanationism.

I then move on to the contrast between modalism and explanationism in the context of applied legal epistemology. By emphasising a few structural connections between knowledge and legal proof, epistemologists have moved quite freely from a modalist account of

knowledge to a modalist account of legal proof. A case in point is Pritchard (2018b), who proposes a modal condition for knowledge (safety) and then extends it to legal standards of proof. On this second issue, I see things differently: in legal epistemology, I endorse explanationism and I reject modalism. Accordingly, I object to influential modal accounts of legal proof and I offer an alternative explanationist account. I develop an account of legal proof as an inference to the best contrastive *de re* why-explanation: by drawing on up-to-date literature on abductive reasoning in the philosophy of science, I bring into focus hitherto unappreciated connections between the nature of legal proof and the structure of scientific explanation. The final upshot is a well-informed account of legal proof which does away with modal conditions and better captures the nature of juridical proof.

Before I turn the attention to kairetic explanationism and to the yet to be developed version of modalism, I must first add another view into the mix (virtue reliabilism) and clarify some key further aspects of modalism and explanationism. In other words, we need to take a closer look. I shall do so in the next chapter.

## 1.6 Chapter Summary

I take myself to have achieved the following in this chapter. Firstly, I have clarified the metaepistemological foundations of questions pertaining to the Gettier problem and anti-Gettier conditions. Secondly, I have laid out the core tenets of the two families of positions that take centre stage in this dissertation: modalism and explanationism. Finally, I've provided a taxonomy of different versions of each family of views.

# Chapter 2

## A Closer Look

### 2.1 Overview of the Chapter

This chapter delves deeper into some versions of modalism and explanationism with a special emphasis on the safety condition for knowledge. By drawing a close comparison with virtue reliabilism, different versions of the safety condition are distinguished and a number of problems for safety are outlined and discussed. It is also shown how kairetic explanationism seems to have the upper hand in handling these two problems. However, appearances are misleading: in the last section I sketch out a strategy to deal with these two objections to the safety: this strategy paves the way for the novel version of safety developed in the subsequent chapters.

### 2.2 Compromising: Virtue-reliabilist Aptness

Recall the key tenets of the two main families of views considered in this dissertation: according to modalist approaches to the analysis of knowledge, knowledge equates with a belief which is true in a relevant set of possible worlds. Champions of the safety condition restrict the set of possible worlds to close and similar possible worlds. Differently, according to explanationist approaches to the analysis of knowledge, knowledge equates with a belief which meets suitably specified explanatory standards. For example, according to kairetic explanationism, to count as knowledge the truth of the belief under consideration must figure essentially in the explanation of why it's hold and pass the so called kairetic test.

We have already focussed on a substantial difference between modalism and explanationism in general and safety and kairetic explanationism in particular. Because the former employs modal notions and the latter does not (remember that explanation is hyperintentional), these two approaches seem to be hard – if not impossible – to reconcile. But once a new player enters the field (virtue reliabilism and aptness theories of knowledge), the differences between modalism and explanationism appear to be less stark. Moreover, and crucially for our purposes, aptness theories of knowledge provide new insights into both

modalism and explanationism: by appealing to aptness, the modalist can develop more precise versions of the safety condition; and by looking at aptness, the explanationist can also clarify the role of explanation in marking a difference between knowledge and ignorance. Accordingly, aptness may be thought to be an attractive compromise between two independently plausible views (safety and kairetic explanationism).

Virtue reliabilist conceptions of aptness are most prominently developed and defended by Ernest Sosa (2007; 2010; 2015; 2021) and John Greco (2010), among others.<sup>1</sup> According to virtue reliabilists (sometimes also called robust virtue epistemology), knowledge is, always and in each case, *success from ability*.<sup>2</sup> More specifically, virtue reliabilist defend different versions of the following theses:

**Performance normativity.** Knowledge is a specific genus of a broader species, i.e. *performances*: knowledge is a type of performance (belief) whose success (truth) is substantially attributable to an agent ability.

A practical analogy with archery usually helps to understand the general structure of performance normativity, and also how it sheds light on the nature of knowledge. Take an archer who skillfully takes a shot and hits the bull's eye: the success of her performance is essentially due and creditable to her competence. Knowledge is just the same – a true belief whose truth is due and creditable to the agent's cognitive ability. This analogy can explain why knowledge is a suitably specified type of success from the exercise of a cognitive ability. In short, knowledge equates with *apt* belief, a belief that is accurate because it's competent (adroit):

**Aptness.** A belief is apt if and only if it is accurate *because* it is adroit (competent).

Two things to note here. First, aptness also provides an elegant anti-Gettier condition. Back to **Sheep in the Field**: even if the agent forms a competent (adroit) true (accurate) belief, such belief is not apt – it's not true *because* it's competent (the truth of the belief is *explained* by the presence of the sheep-looking dog nearby).<sup>3</sup> Secondly, cognitive abilities are standardly conceived as *dispositions* to reliably form true beliefs, and the *because/due to* relation lends itself to two distinct readings:

1 **Explanatory-salience aptness.** S's belief that *p* is apt if and only if the most salient part of a causal explanation for why S's belief that *p* is true is S's stable disposition to believe truly. (Greco 2010)

---

<sup>1</sup>See also Kelp (2011, 2013, 2018), Miracchi (2015), Turri (2016), Riggs (2009), and Carter (2016) for further (and different!) defenses of virtue reliabilism. See Turri & Greco (2017) and Greco & Kelp (2020) for an up to date overview of virtue reliabilism.

<sup>2</sup>The slogan is due to John Greco.

<sup>3</sup>The explanatory power of aptness may come at a high price. As some authors have influentially argued (Lackey 2007, 2009; Pritchard 2010, 2012a), aptness-based theories of knowledge struggle to accommodate common instances of testimonial knowledge. For the purposes of this chapter, I shall set aside problems pertaining to aptness and testimony.

2 **Ability-manifestation aptness.** S's belief that  $p$  is apt if and only if S's belief that  $p$  manifests a *complete competence* of the agent. (Sosa 2015)

These two readings of aptness help to develop more fine-grained versions of modalism and explanationism. But before moving on to such versions, it's worth taking a closer look at the similarities between aptness and modalism on the one hand and aptness and explanationism on the other.

## 2.3 Between Modalism and Explanationism

The main point of agreement between advocates of safety and advocates of aptness lies in the modal conception of abilities defended by virtue reliabilists. Recall that according to both Greco and Sosa, abilities are reliable *dispositions* to form true beliefs. Crucially, dispositions naturally lend themselves to a *modal counterfactual* reading. To say that object O has a disposition to  $\phi$  in condition C is to say that if O were in condition C, then O would (likely)  $\phi$ . This is a trigger-manifestation *conditional*: if object O was in conditions C, would O likely to  $\phi$ ?<sup>4</sup> The (more fundamental) metaphysics of dispositions clarifies the nature of abilities, and we can ask the same trigger-manifestation conditionals to investigate more specific *cognitive* abilities: if S formed a belief in conditions C, would S's belief (likely) be true?

Crucially, a dispositional conception of abilities yields a modal reliability condition that nicely captures the spirit of safety.<sup>5</sup> It's not so hard to see why: if knowledge is success from ability, then the agent continues to form true beliefs in the close possible worlds where they continue to exercise their cognitive abilities (which are, recall, modal dispositions to form true beliefs). Accordingly, by specifying the conditions relative to which a cognitive ability is successfully and completely exercised, we can get a better purchase on the details of a safety condition on knowledge.

A different but relevant point of agreement between aptness anti-Gettier conditions and explanationism can be found in Greco's *explanatory salience* conception of aptness. According to Greco, knowledge is an achievement in the sense that it's a success *causally explained* by cognitive abilities. For this reason, Greco opts for an explanatory reading of the because relation featuring in the formulation of aptness. To characterise this explanatory reading, Greco offers the following remarks, worth quoting in full:

What factors govern the distribution of explanatory salience? That is not well understood, but we can say some things about it. *First, explanatory salience is often governed by relevant interests and purposes [Feature 1].* Thus one

---

<sup>4</sup>See Sosa (2015: 96) and Kelp (2018: 20) for a thorough exposition of the analogies between cognitive abilities and dispositions. Vetter (2014, 2015) develops influential arguments against a modal and dispositional conception of abilities.

<sup>5</sup>Kelp (2018: 21) makes this point as follows: "It may be worth noting that the trigger-manifestation conditionals effectively impose a reliability condition on abilities. Abilities are, by nature, reliably connected to the relevant successes in the sense at is specified by the trigger-manifestation conditional." The relevant sense of the reliability under consideration is modal rather than, say, merely probabilistic.

reason we look to sparks rather than oxygen to explain the warehouse fire is that we more easily control the presence of sparks than the presence of oxygen. Likewise, we readily attribute a crash to the drunk driver, presumably because we have an interest in regulating drunk driving. *Second, explanatory salience attaches to the odd or unusual [Feature 2]*. Thus suppose that a fire starts in an area where there is usually no oxygen, perhaps in some special space in a chemical lab. In that case, we might very well attribute the fire to the presence of oxygen.<sup>6</sup> (Greco 2012a: 10)

From this, we can highlight two main features of explanatory salience:

- 1 *Sensitivity to practical interests.*
- 2 *Relevance of abnormal factors.*

Explanatory salience is constrained by the practical purposes and interests underlying the candidate explanations. In the case of knowledge, we will therefore focus more on how the exercise of abilities may or may not explain the truth of the belief under consideration. Moreover, our attention will be also directed to abnormal features of the case. In Kelp's nice terminology, abnormalities are 'salience magnets' (Kelp 2018: 38). For instance, in **Sheep in the field**, the fact that there is a sheep that looks like a dog is a type of abnormal feature that becomes relevant for the explanation of the truth of the belief under consideration.

These two features of explanatory salience can be incorporated in any explanationist account of knowledge. For example, explanationists can characterise their notion of explanation as the best explanation *given contextual practical purposes*.<sup>7</sup> Moreover, explanationists can also stipulate that the explanation of why S believes that *p* should not feature abnormal factors. Overall, Greco's explanatory salience reading of the aptness condition displays interesting point of contact with explanationist approaches to the analysis of knowledge.

## 2.4 Fine-grained Versions of Safety

With the notion of aptness on the table, we can now move on to develop three distinct (and more precise) versions of the safety condition. The first relativises safety to belief-forming methods; the second and the third unpack safety in terms of aptness by appealing to Sosa's and Greco's notions. Overall, each of these versions advances our understanding of the safety condition on knowledge.

---

<sup>6</sup>Italics are mine.

<sup>7</sup>Contrastivist versions of explanationism would arguably fit the bill.

### 2.4.1 Standard Method-relative Safety

The key insight of a safety condition for knowledge is that a safe belief can't easily be false: in the close possible worlds where the subject forms the belief under consideration, such belief continues to be true. One immediate question pertains how *tolerant* a safety condition can be. This question translates into how to carve out the relevant range of possible worlds: can the belief be false in one or more close worlds, or does it have to be true in *every* close world? The first choice point to elaborate on a safety condition is between *strong* and *weak* formulations of safety:

**Strong Safety.** S's belief that *p* is safe if and only if S's belief that *p* is true in the actual world and in *every* close possible world in which S believes that *p*, *p* is true. (Greco 2012b: 196)

A more tolerant safety condition admits false beliefs in some close worlds:

**Weak Safety.** S's belief that *p* is safe if and only if S's belief that *p* is true in the actual world and in *most* close possible worlds in which S believes that *p*, *p* is true. (Greco 2012b: 196)

So much for the tolerance of the safety condition. But of course this can't be the full story: absent any further specifications, a rough and unrestricted safety condition invites counterexamples like the following:

**Grandmother.** A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset. Yet this does not mean she doesn't know he is well (or at least ambulatory) when she sees him. (Nozick 1981: 179)

**Dachshund.** I look at a nearby dachshund and truly believe that what I see is a dog. Had I not been seeing a dachshund, I would have been seeing a wolf, and would have falsely believed myself to be seeing a dog. (Goldman 1976)

These counterexamples were originally raised against a sensitivity condition, but they also apply against a safety condition. The point of these counterexamples is the following: while in close possible worlds the agents seem to form false beliefs (because the grandson is ill, or the dog is a wolf), they nevertheless seem to know in the actual world. This led Nozick (and other modalists) to relativise the safety (or sensitivity) condition to the agent's evidential situation and to the belief-forming methods under consideration (also dubbed bases. Commenting on another counterexample raised against safety, Pritchard emphasises the importance of keeping the belief-method under consideration fixed as follows:

The crux of the matter is that we need to keep the subject's actual evidential basis fixed, and of course his actual evidential basis for his belief is formed by

consulting the reliable and unaffected clock. While it is lucky that the subject has this evidential basis (in that there are close possible worlds where it is absent), it is not lucky that she forms a true belief on this basis. Indeed, in all close possible worlds where she continues to enjoy the same evidential basis she continues to form a true belief.<sup>8</sup> (Pritchard 2018a: 9)

The focus on the belief-forming methods helps to make progress on the counterexamples raised above: what matters for the safety condition is that the belief under consideration continues to be true *given the way it was formed*. But how are we to individuate belief-forming methods in a coherent, informative and principled way? On this point, matters become more complicated. Modalists like safety and sensitivity theorists standardly hold that belief-forming methods can be sorted according to the following criteria:

1 *Internal vs external* individuation of bases.<sup>9</sup>

2 *Coarse-grained vs fine-grained* individuation of bases.<sup>10</sup>

If the criterion is internal, then belief-forming methods are restricted to the agent's mental states (e.g., visual appearances such as *seeming to see a dog*: crucially, mental content is individuated narrowly). External criteria deny this, and build external elements into the individuation of belief-forming methods (e.g., *seeing an actual dog*). A different degree of specificity separates coarse-grained from fine-grained criteria of individuation of belief-forming methods: the former are less specified (e.g., seeing a dog *tout court*), the latter make reference to additional information (seeing a dog *in normal environmental conditions*).<sup>11</sup>

Those tasked with identifying a successful criterion of individuation of methods face the following challenges. The first challenge is to provide a criterion which is principled, informative and that avoids any charge of *ad hocness*. The second challenge is to come up with a sufficiently high degree of specificity for the belief-forming methods. This is a type of *generality problem*.<sup>12</sup> Because belief-forming methods can be specified more or less precisely (*seeing a dog vs seeing a dog on the 24th of April 2022 in normal visual conditions*), the degree of specification can make a difference between knowledge and ignorance; moreover, the degree of specification can be crucial in assessing the modal profile of the belief-forming method across relevantly close similar possible worlds.<sup>13</sup>

<sup>8</sup>Emphasis mine; I return to said counterexample in more detail in the next chapter(s).

<sup>9</sup>This terminology is employed by both Nozick (1981) and Pritchard (2005, 2018a).

<sup>10</sup>The terminology is employed by Williamson (2009), Goldman (2009), Alfano (2009), Becker (2012) and Rabinowitz (2011).

<sup>11</sup>Following Alfano (2009), a logical criterion for sorting belief-forming methods reads as follow:

$$(\forall m_1)(\forall m_2)[R_M(m_1, m_2) \equiv (m_1 \in M \ \& \ m_2 \in M)],$$

. As Alfano clarifies,  $m_1$  and  $m_2$  are tokens of belief-forming methods and  $M$  is a method type.

<sup>12</sup>An early version of the problem is already discussed in Goldman (1979, though see Conee & Feldman (1998) for a *locus classicus*, Heller (1995) for an early solution and Beddor & Goldman (2021: §2.3) for recent a survey of possible solutions to the generality problem.

<sup>13</sup>The generality problem for belief-forming methods is similar but distinct from the generality problem affecting process-realist theories of justification (Lyons 2019: 466, fn. 4). For a thorough discussion of the differences between these two types of generality problems, see Lyons (2021: §3).

One good way to make progress on these challenges is to specify belief-forming methods with references to aptness and cognitive abilities.

### 2.4.2 Ability-relative SSS Safety

Given that there are good reasons to relativise safety to belief-forming methods or bases, it is open to the advocate of modalism to appeal to virtue-theoretic criteria of individuation of belief-forming methods. As mentioned above, virtue reliabilist insist on the key role of dispositional cognitive abilities for their theory of knowledge. A promising strategy here is to understand methods and the way beliefs are formed in reference to cognitive abilities. Recall Pritchard's formulation of safety:

**Pritchard's Safety.** S's belief is safe if and only if in most nearby possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, and in all very close nearby possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world, the belief continues to be true. (Pritchard 2005, 2007)

The move consists in swapping “the way the belief is formed” with a virtue reliabilist competence. More precisely, competences are understood along the lines of the following tuple:

- The innermost seat (Se), located in the agent's brain (also called “skill”).
- The inner shape (Sh), corresponding to the agent's general condition.
- The external situation (Si), namely the environment in which the agent is performing.

Sosa provides the following example to illustrate the structure of competences:

With regard to one's competence in driving, for example, we can distinguish between (a) the innermost driving competence that is seated in one's brain, nervous system, and body, which one retains even while asleep or drunk; (b) a fuller inner competence, which requires also that one be in proper shape, that is, awake, sober, alert, and so on; and (c) complete competence or ability to drive well and safely (on a given road or in a certain area), which requires also that one be well situated, with appropriate road conditions pertaining to the surface, the lighting, etc. The complete competence is thus an SSS (or an SeShSi) competence.<sup>14</sup> (Sosa 2017: 191, 192)

The SSS tuple applies to virtually any other competence: from playing a sport or an instrument including also belief-forming methods. Accordingly, we can restrict belief-forming methods to the exercise of a skill in a suitable shape and situation:

---

<sup>14</sup>See also Carter (2021: §4) for an illuminating discussion of Sosa's notion of abilities.

**SSS-Safety (Sosa).** S belief that  $p$  is safe if and only if in all close possible worlds in which S continues to form her belief about the target proposition exercising the same skill (Se) while in the same shape (Sh) and external situation (Si), S's belief is true.<sup>15</sup>

Take for instance any knowledgeable perceptual belief: once we keep the skill, shape and situation fixed, these beliefs will continue to be true in relevantly closed possible worlds. As long as the agent exercises their perceptual skill in an appropriate shape and in a suitable situation, these beliefs will be modally robust. The same story can be told about other types of beliefs based on memory, testimony and inferences.

This is a type of *restricted safety*: rather than requiring one's belief to be true across possible worlds *simpliciter*, it focusses on a specific (and restricted) set of circumstances. There are important choice points for picking out the relevant set of circumstances, the range of propositions and the degree of reliability; Virtue reliabilists have a more internal disagreement on how to best do so. As we saw above, Sosa applies a theory of dispositions to inform fill the parameters of Seat, Shape and Situation. Similarly, Greco (2010, 2012a, 2020b) thinks of abilities as relative to environments and he restricts safety to the exercise of abilities in "appropriate" conditions. He further thinks of conditions in terms of *modal environments* – swaths of possible worlds centred around the actual world. When it comes to range of propositions and degrees of reliability, he appeals to *practical takes*: abilities are relative to modal environments and information-dependent practical tasks that inform range of propositions and degree of reliability. Relative to that specific modal environment, abilities will yield safe beliefs. This is Greco's full account of abilities:

S has a knowledge-producing ability A(Se/Sh/Si/R/D) relative to a modal environment E if and only if: S has a disposition seated in inner seat Se to believe truths in an appropriate range R of propositions when in appropriate Sh and appropriate Si, and in E, with appropriate degree of reliability D. (Greco 2020b: 133)

And in turn, such an account entails the following restricted version of safety:

**SSS-Safety (Greco).** S's belief that  $p$  is SSS-safe: for every proposition in range R, not easily would S believe  $p$  (while retaining Se and in Sh, Si) when  $p$  is false. (Greco 2020b: 134)

Once belief-forming methods are individuated with reference to abilities, the safety condition becomes clearer and more informative. Indeed, this virtue-theoretic criterion of individuation of belief-forming methods meets the challenges outlined in the previous section. Thanks to the reference to the SSS tuple, it is informative and it avoids charges of *ad hocness*. And by filling the details in light of skill, shape and situation of the agent, it also achieves a sufficiently high level of specificity: the SSS tuple allows us to make

---

<sup>15</sup>This formulation is borrowed from both Greco (2020a: 5152) and Carter (forthcoming: Chapter 2) presentations.

the right predictions when it comes to evaluate whether the belief under consideration amounts to knowledge.

## 2.5 Problems with Safety (and Explanationist Solutions)

Advocates of a safety condition for knowledge insist on the importance of the *modal robustness* of belief-forming methods. This in turn raises important questions pertaining to the *criteria* of individuation of belief-forming methods. One way to make progress on these questions is to adopt a virtue-theoretic criterion, and unpack belief-forming methods in terms of dispositional cognitive abilities – and with reference to the SSS tuple. The result is a full-fledged safety condition, which (at the very least) can be developed and formulated in a plausible way.

Of course, being in a position to *articulate* the safety condition properly doesn't mean that we should *accept* it. Indeed, there are two big persistent problems for any safety condition for knowledge – problems that highlight its inadequacy as anti-Gettier condition or condition for knowledge. Let me now outline each in turn.

Firstly, there are cases of unsafe knowledge in which the agent forms a false belief in a relevantly similar close possible world and yet they intuitively know in the actual world. Kelp (2009, 2016) offers the following one:

**Frankfurt Clock.** Russell's arch-nemesis, a powerful demon, has an interest that Russell forms a belief that it is 8:22 by looking at the grandfather clock in the hallway when he comes down the stairs. Russell's arch-nemesis is prepared to do whatever it may take in order to ensure that Russell acquires a belief that it is 8:22 by looking at the grandfather clock when he comes down the stairs. However, Russell's arch-nemesis is also lazy. He will act only if Russell does not come down the stairs at 8:22 of his own accord. Suppose, as it so happens, Russell does come down the stairs at 8:22. Russell's arch-nemesis remains inactive. Russell forms a belief that it is 8:22. It is 8:22. The grandfather clock is working reliably as always.

The main point of the case is that, *despite* the close possibility of error, Russell does nevertheless come to know the time. Had Russell come down the stairs a few seconds earlier and formed the target belief in the same way, he'd have formed a false belief. As such, safety doesn't seem to be necessary for knowledge: a true belief can easily be false *and yet* constitute knowledge. More such cases can be generated, and they all share the same structure: even with a modally close possibility of error, the agents can come to know the target proposition.<sup>16</sup>

Once this type of cases are raised *in conjunction* with fake-barn style cases, the safety theorist faces the following dilemma:

---

<sup>16</sup>Comesaña (2005), Baumann (2008 and 2014) also develop similar cases.

- (*horn 1*) Either safety is going to be too strong to predict knowledge in Frankfurt-style cases
- (*horn 2*) or it's going to be too weak to predict absence of knowledge in barn-style cases.

No matter which horn is picked, safety and knowledge come apart. As Kelp puts it, this is the safety dilemma that puts pressure on the tenability of safety as anti-Gettier or necessary condition on knowledge.

Yet more cases question the *sufficiency* of safety for knowledge. For example:

**TrueTemp.** Temp forms beliefs about the temperature by consulting a broken thermometer. However, Temp has a guardian angel in the room who controls the thermostat, ensuring that the room's temperature matches the reading displayed on the thermometer. Assuming the angel manipulates the thermostat in all nearby worlds, any belief Temp forms about the temperature in such worlds will be true. Pritchard: (2012a)

**Necessary Truths.** Emma is using a calculator to count the sum of  $12 \times 13$ . As a result, she forms a true belief that  $12 \times 13 = 156$ . However, the calculator is actually broken and it's generating answers at random. Emma's belief is true in (every) close possible worlds. (Pritchard: 2012a)

The beliefs under consideration are true in the actual world and also true in close (e.g. TrueTemp) or every possible worlds (necessary truths), yet they don't amount to knowledge. Even worse for the safety theorists who are concerned with finding an anti-luck condition, these beliefs are safe and yet only luckily true (they are veritically lucky).

Taken together, these cases suggest that the safety condition doesn't seem to be neither necessary nor sufficient for knowledge.

Interestingly, explanationism seems to succeed exactly where safety fails. To illustrate this point, let's pause to appreciate the reason why safety is neither necessary nor sufficient for knowledge in the cases discussed in the previous section. Taken for instance **Frankfurt Clock**: as some authors have noted (Bogardus 2014), the belief under consideration meets some key explanatory connection: because the clock is functioning properly and reads correctly, the agent forms a correct belief about the time. This is precisely why the modal proximity of error doesn't prevent the agent from acquiring knowledge: explanationists have an easy time to avoid the safety dilemma.

A similar story can be told about counterexamples to the sufficiency of safety for knowledge. Despite the modal robustness, the beliefs under consideration in **TrueTemp** and **Necessary Truths** fail to meet the right explanatory connection: Temp's belief is true because there's a guardian angel and the calculator is yielding those results because it's malfunctioning. Again, lack of explanatory connections explain lack of knowledge.

The point about explanatory connections can be made in many ways depending on the relevant notion of explanation at play. One might resort to Strevens' kairetic test, and

hold that the truth of the belief figures in the complete explanation of the beliefs under consideration. Perhaps the truth of the belief under consideration is the *best* explanation of why the agent holds the belief. Or, by borrowing Greco’s terminology, the truth of the belief under consideration is a salient explanation of why it’s held. Either way, it’s clear that explanationism (no matter how unpacked) has a substantial advantage when it comes to deal with the counterexamples raised against the necessity and sufficiency of safety for knowledge.

How can the safety theorists respond? A key upshot of this discussion is the following: the safety condition ought to be made more precise. Take for instance Pritchard’s formulation. In his formulation, it’s clear that too many details remain to be filled: the relevant external circumstances, belief-forming methods, close possible worlds, and range of propositions. Virtue reliabilist have the resources to do so, and can derive a *restricted* safety condition from their preferred notion of ability.

Should the safety theorists retreat and adopt a version of virtue reliabilism? Would this help them to deal with the counterexamples to safety? Matters are complicated, and a lot will depend on which version of virtue reliabilism is picked and endorsed. The safety theorist may restrict belief-forming methods to Greco’s notion of abilities and deal with counterexamples like **Necessary Truths** and **TrueTemp** like explanationists do. But Sosa’s manifestation readings of abilities may be less helpful: for example, TrueTemp’s plausibly has an ability to form belief truly. Their belief-forming method passes the trigger-manifestation conditional test: if S formed a belief in conditions C, would S’s belief (likely) be true? If Temp formed a belief in these favourable conditions, their beliefs is guaranteed to be true. Perhaps unsurprisingly, this led Sosa to accept an (animal) knowledge verdict in the case of TrueTemp (Sosa: 2010, 2015). Accordingly, retreating to virtue reliabilism may help with *some* though not *all* the objections to safety.

We can sum up the comparison between a specific version of modalism (safety) and explanationism with the following table. The scoreboard reads thusly:

	Safety Dilemma	Necessary Truths	TrueTemp
Unrestricted Safety	x	x	x
Explanationism	v	v	v
Restricted Sosa SSS-Safety	?	v	x
Restricted Greco SSS-Safety	?	v	v

## 2.6 The Way Forward

At this point, it’s worth pausing to set the record straight about the dialectic considered so far. The chief version of modalism, safety, equates knowledge with a modally robust true beliefs: this equation requires a suitable specification of belief-forming methods, and a good way to specify belief-forming methods is to appeal to the SSS tuple endorsed by virtue reliabilist accounts of knowledge (e.g., Sosa’s aptness). While this helps to

articulate the safety condition well, there are remaining problems with the necessity and sufficiency of safety for knowledge – problems that are made vivid by the safety dilemma and other counterexamples.

What about explanationism? Different versions of explanationism seem to succeed precisely where safety fails, and the reason why they do so is the insistence on explanatory connections that are met in the case of knowledge and not met otherwise. These claimed advantages of explanationism seem to suggest that the right anti-Gettier condition is hyperintensional and explanatory rather than essentially modal.

The following chapters will question this assumption. Chapter 3 provides a solution to the safety dilemma by indexing safety to belief-forming methods and environments. Chapter 4 reveals the inadequacies of one recent version of explanationism – kairetic explanationism – as anti-Gettier condition. Chapter 5 makes the index to belief-forming methods and environments more precise by appeal to etiological proper function; this novel version of safety deals with each counterexample raised against in the literature.

## **2.7 Chapter Summary**

In this chapter I've taken a closer look at modalism and explanationism through the lens of virtue reliabilism. This comparison brought up important similarities: the explanatory salience version of aptness is close enough to explanationism, and the notion of abilities defended by virtue reliabilists helps to better unpack the safety condition by providing a virtue-theoretic criterion of individuation of belief-forming methods. I also outlined some problems for the safety condition that questions its necessity and sufficiency for knowledge, and I've shown how explanationism seems to have the upper hand in handling these two problems.

# Chapter 3

## Solving the Safety Dilemma

### 3.1 Overview of the Chapter

Despite the substantial appeal of the safety condition, Kelp (2009, 2016, 2018) has raised a difficult challenge for safety-theoretic accounts of knowledge. By combining Gettier-style fake barn cases with epistemic Frankfurt cases, he concludes that no formulation of safety can be strong enough to predict ignorance in the former and weak enough to accommodate knowledge in the latter. In this chapter, my contribution is two-fold. Firstly, I take up Kelp's challenge and I show that, once properly understood, safety successfully rises to it. Secondly, I draw a more general lesson on the safety condition: a satisfactory solution to Kelp's challenge calls for a revision of the standard formulation of safety, which must be indexed to *both* methods and environments. My conclusion will be that an environment-relative version of safety not only meets Kelp's challenge, but it also advances our understanding of the safety condition on knowledge.

### 3.2 The Hardest Challenge for Safety

Recall that, according to the safety condition, knowledge requires protection from error: if one knows, then one could not have *easily* been wrong in similar cases. As a modalist condition, safety is formulated by reference to possible worlds and indexed to methods of belief formation:

**Safety.** In most or all close possible worlds in which S believes that  $p$  via the same method of belief formation  $M$  that S uses in the actual world,  $p$  is true.<sup>1</sup>

As we have seen, safety has considerable explanatory power. As an anti-luck condition, it promises to solve the Gettier problem: safety delivers the correct verdict of ignorance in various Gettier-style cases. By denying knowledge of lottery propositions, safety deals

---

<sup>1</sup>Prominent defenders of safety include Sainsbury (1997), Sosa (1999), Williamson (2000), and Pritchard (2005, 2007). For recent discussion of the safety condition on knowledge, see Bernecker (2020), Greco (2020a) and Hirvelä (2019). See Rabinowitz (2011) for an overview.

also with another much debated issue in epistemology – the lottery problem.<sup>2</sup> Moreover, safety complies with the closure principle, and, unlike sensitivity, it does not open the door to a disturbing sceptical threat (cf. Pritchard 2008: 446).

Despite such substantial appeal, the safety condition has now fallen on hard times. As we have seen in the previous chapter, in various places Kelp (2009, 2016, 2018) has challenged safety with a troubling conjunction of cases: Gettier-style fake barn cases and more ingenious epistemic Frankfurt cases. Begin with the former:

**Fake barn.** Barney, a reliable barn spotter, is driving through the countryside. He looks out of the window, sees a barn and comes to believe that he is looking at a barn. Whilst Barney's belief is true, unbeknownst to him, the structure he is looking at is the only real barn in an area filled with fake barns that are indistinguishable from real barns. (cf. Goldman 1976)

Intuitively, Barney does not know that he's looking at a barn, and safety neatly explains why. Barney's belief could too easily have been false: in many similar close possible worlds where he forms the same belief via the same method (i.e., visual perception), Barney's belief is false because he's looking at one of the many papier-mâché copies of the real barn. Here, safety delivers the correct verdict. But now consider the following apparently similar case:

**Frankfurt Clock.** Russell's arch-nemesis, a powerful demon, has an interest that Russell forms a belief that it is 8:22 by looking at the grandfather clock in the hallway when he comes down the stairs. Russell's arch-nemesis is prepared to do whatever it may take in order to ensure that Russell acquires a belief that it is 8:22 by looking at the grandfather clock when he comes down the stairs. However, Russell's arch-nemesis is also lazy. He will act only if Russell does not come down the stairs at 8:22 of his own accord. Suppose, as it so happens, Russell does come down the stairs at 8:22. Russell's arch-nemesis remains inactive. Russell forms a belief that it is 8:22. It is 8:22. The grandfather clock is working reliably as always. (Kelp 2016: 28)

Intuitively, Russell does know the time. Yet, his belief appears to be unsafe: while Russell's belief is true, it could too easily have been false. For sake of vividness, consider the possible worlds where Russell comes down the stairs a few minutes earlier. In these nearby worlds, Russell forms a false belief via the same method, which Kelp identifies as Russell's ability to read clocks (Kelp 2016: 28). **Frankfurt Clock** is offered as a straightforward instance of unsafe knowledge: Russell knows, but the belief in question is unsafe. In this case, safety delivers the wrong verdict.

---

<sup>2</sup>Consider Chisholm's sheep in the field Gettier-style case: in nearby worlds where there's no sheep behind the rock, the agent forms a false belief, hence no knowledge. Same with lottery-style propositions: since there's a similar nearby world where your ticket wins, your belief that your ticket is a loser could too easily be false (Pritchard 2012a: 253-254).

Kelp's argument casts into serious doubts the success of safety as anti-Gettier condition on knowledge. To fully appreciate the significance of his challenge, we must consider **Fake Barn** in conjunction with **Frankfurt Clock**.<sup>3</sup> It looks as though no version of the safety principle will be strong enough to predict ignorance in Gettier-style cases like **Fake Barn** and, at the same time, weak enough to accommodate knowledge in **Frankfurt Clock**.<sup>4</sup> Kelp calls this "the safety dilemma". He draws the following pessimistic conclusion, worth quoting in full:

I'd like to add that meeting this challenge is by no means trivial. Most standard proposals in the literature are prone to fail it in one way or another. Modal accounts, such as safety and sensitivity-based accounts, fail it because the conditions they impose are too strong: they predict absence of knowledge in Frankfurt cases. Other accounts fail in that the proposed conditions are too weak to explain the absence of knowledge in fake barn cases. If there exists a workable account of knowledge in terms of justified true belief and an anti-Gettier condition that gets both fake barn cases and Frankfurt cases right, epistemology still awaits its discovery. (Kelp 2016: 36)

Safety theorists have not met Kelp's challenge yet, at least not in Kelp's terms. To wit, Pritchard (2009: Section 3) tentatively defends a no-knowledge verdict in **Frankfurt Clock**. Given the presence of environmental luck in both **Fake Barn** and **Frankfurt Clock**, he maintains that neither is a genuine case of knowledge. However, Pritchard's response is counterintuitive: the knowledge intuition in **Frankfurt Clock** is plausible, and it's hardly explained away by considerations pertaining to environmental luck. Other safety theorists (e.g., Grundmann 2018) have attempted to capture the knowledge intuition in **Frankfurt Clock** by appealing to a more fine-grained individuation of belief-forming methods. As we will see, this strategy struggles to explain absence of knowledge in **Fake Barn**. At present, Kelp's challenge still stands.

In this chapter, I revisit Kelp's challenge and develop a better solution than the ones currently on offer. I shall argue that a properly understood safety condition successfully captures the intuitions of knowledge and ignorance in **Frankfurt Clock** and **Fake Barn**. The two vignettes display an important but hitherto overlooked difference in their respective environments. Once brought into focus, such difference has far-reaching consequences. Firstly, it provides a principled way out of Kelp's dilemma, and, in doing so, it also weakens one of the main motivations for the recent knowledge-first shift in the virtue-epistemological camp (e.g., Miracchi 2015; Kelp 2018). Crucially, Kelp raises

---

<sup>3</sup>It's worth add a point of clarification on the challenge raised by Kelp. Kelp's epistemic Frankfurt case closely resembles other counterexamples to the necessity of safety for knowledge moved by Neta and Rohrbaugh (2004), Comesaña (2005), Baumann (2008) and Bogardus (2014). However, Kelp takes these counterexamples one step further by conjoining them with fake-barn cases, thus posing an even more difficult challenge for the safety condition on knowledge.

<sup>4</sup>For the purpose of this chapter, I assume the no-knowledge verdict in **Fake barn** to be correct. However, I also note that both epistemologists (Lycan 2006; Sosa 2010: 472-473) and experimental philosophers (Colaço et alia 2014) have adduced arguments to challenge the correctness of the no-knowledge verdict in cases like **Fake Barn**.

this dilemma to motivate the move from traditional (or indirect) virtue epistemology to knowledge-first (or direct) virtue epistemology: with a principled way out of his dilemma on the table, knowledge-first virtue epistemology loses one of its main motivations.<sup>5</sup> Secondly, this key environmental difference is of independent interest to safety theorists: in fact, it further elucidates the safety condition on knowledge, which is best understood as relative not only to belief-forming methods but also to environments.

Having clarified my aims, I now proceed to criticise the two extant solutions to Kelp's dilemma. The former is conciliatory: by adopting a more fine-grained individuation of belief-forming methods, it attempts to reconcile safety with the knowledge verdict in **Frankfurt Clock**. The latter is revisionary: by insisting on the presence of environmental luck, it defends a verdict of ignorance in both **Frankfurt Clock** and **Fake Barn**. Neither is ultimately successful. To do better, I close by outlining a refined environment-relative safety condition: unlike its conciliatory and revisionary rivals, this solution rises to Kelp's challenge.

### 3.3 Belief-forming Methods

To avoid counterexamples like Nozick's grandmother case (Nozick 1981: 179) or Goldman's dachshund vignette (Goldman 2009: 80), modal conditions on knowledge are usually indexed to belief-forming methods.<sup>6</sup>

Let's pause to appreciate the strength and the point of these counterexamples. In Nozick's vignette, the agent forms a false belief in nearby worlds, but the belief-forming method is crucially different (Granny is told that her grandson is well instead of seeing that he's well). These cases have led safety theorists to focus not only on the truth value of the target proposition in nearby worlds, but also on whether the agent's belief continues to be true in nearby worlds given the way it was formed in the actual world. For further discussion, see Pritchard (2008: 444-446; 2013: 158) and Rabinowitz (2011: Section 3).

So much for belief-forming methods. But could the safety theorist also meet Kelp's challenge by appeal to a more specific, fine-grained individuation of methods? Unfortunately, the answer is negative. In this section, I argue that while initially promising, this conciliatory strategy ultimately fails. In fact, by yielding a knowledge verdict in both **Frankfurt Clock** and **Fake Barn**, the conciliatory strategy simply doesn't distinguish between our two cases of interest.

Let's focus on methods. In assessing **Frankfurt Clock**, Kelp comments as follows:

Intuitively, Russell knows that it is 8:22 in this case. After all, we may assume,

---

<sup>5</sup>See Kelp (2016; 2016: 35-36; 2017: 240-243; 2018: 78-79) for the connection of the safety dilemma with knowledge-first virtue epistemology. The other main motivation for knowledge-first virtue epistemology stems from the difficulties of traditional virtue epistemology in dealing with Gettier cases (Miracchi 2015; Kelp 2017: 229-230; Kelp 2018: 47-52). However, Hirvelä (2019) argues that, pace Kelp and Miracchi, traditional virtue epistemology can in fact handle these Gettier cases successfully. Special thanks to an anonymous reviewer of *Synthese* for encouraging me to emphasise this point

<sup>6</sup>To be precise, Nozick's and Goldman's cases are offered in support of method-relative sensitivity rather than safety. However, similar considerations equally motivate a method-relative version of safety.

he has the ability to read the clock and forms a belief via an exercise of this ability. *Moreover, the clock is actually functioning properly, and the reading is accurate.* At the same time, Russell's belief is not safe. After all, among the very close nearby worlds are worlds at which Russell comes down the stairs a minute earlier or later. *At those worlds, the demon intervenes with the result that Russell takes a reading from an inaccurate clock and ends up with a false belief that it is 8:22. At the same time, Russell continues to form his belief in the same way as in the actual world, that is, by exercising his clock-reading ability.* Since safety does not tolerate false beliefs at very close nearby possible worlds at which the agent forms his belief in the same way as in the actual world, Russell's belief turns out unsafe. (Kelp 2016: 28; emphasis mine)

As the quote suggests, Kelp identifies the relevant belief-forming method as the general ability to read a clock. To make his case, he moves from worlds where the clock is functioning properly and accurately to worlds where the clock is *not* functioning properly and it's *inaccurate*. However, in such worlds the belief-forming method is crucially different. One thing is to form beliefs on the basis of a properly functioning accurate clock; another is to form beliefs on the basis of a manipulated, inaccurate clock. If in lieu of the latter we assume the former belief-forming method and we then focus on the worlds where the clock is functioning properly, Russell does keep forming true beliefs. Once specified as the more fine-grained ability to read a *properly* functioning clock, the belief-forming method turns out safe: in this way, safety captures the knowledge verdict in **Frankfurt Clock**.<sup>7</sup>

In fairness to the conciliatory strategy, the focus on worlds where the clock is functioning properly enjoys independent plausibility. If we agree with Kelp that the relevant belief-forming method is Russell's ability to read a clock, it makes sense to specify the relevant ability by including the fact that the clock is functioning properly. This is clearer if we think of abilities along the lines of Sosa's influential triple-S model (Sosa 2010: 465-467; 2017: 191-192) that we already outlined in the previous chapter. For every ability, we can specify the innermost seat (or skill) in the agent's brain, nervous system and body, the agent's inner shape (are they drunk, sober, asleep, alert?) and the external situation the agent is in (are there any relevant external interferences or intervening factors?). Russell gains knowledge only through the exercise of a complete ability to read a clock: the complete ability is exercised only when each S is in play, that is, only when he is properly situated. The relevant situation here must include a properly functioning clock: in the worlds where Russell forms a false belief, the relevant ability is different due to the crucial change of situation (the clock stops). Once the belief-forming method is specified by reference to Russell's complete ability, his belief is safe.<sup>8</sup>

<sup>7</sup>See Grundmann (2018: 10-12) for a brief defence of this line of response Kelp's epistemic Frankfurt case.

<sup>8</sup>Hirvelä (2019: 1178-1180) also defends a similar virtue-theoretic approach to individuate safe belief-forming methods.

While *prima facie* promising, the conciliatory strategy is ultimately unsuccessful. Recall the key challenge of the safety dilemma: to accommodate presence of knowledge in **Frankfurt Clock** and to capture the ignorance intuition in **Fake Barn**. A more fine-grained individuation of methods achieves the former but crucially fails to deliver on the latter. To see why, let's consider again **Fake Barn**. If we include the fact that the clock is functioning properly in **Frankfurt Clock**, by parity of reasoning we will also have to concede that Barney is looking at the only real barn in **Fake Barn**. Advocates of the conciliatory strategy will have to individuate a more fine-grained belief-forming method in both cases: so, just like Russell reads from a properly functioning clock instead of merely reading from a clock, Barney also sees the real barn instead of merely having a visual appearance of a barn.<sup>9</sup> But, once specified in the same way as Russell's, Barney's belief will be safe too: if he forms beliefs by looking at *that* real barn, then he will trivially form true beliefs in every close world in which he's looking at *that* very same barn. Otherwise put, this more specific belief-forming method yields a *de re* true belief in every close world.<sup>10</sup> A more fine-grained individuation of methods may reconcile safety with the knowledge verdict in **Frankfurt Clock**, but it inevitably fails to explain ignorance in **Fake Barn**.

This more fine-grained criterion of individuation of belief-forming methods spells trouble for the safety theorist because it's so specific to be factive, and thus yields only true beliefs: the clock is functioning properly and accurately; Barney sees that there's a barn and in fact there is a barn. I am not the first to raise this type of worries concerning such way of individuating belief-forming methods. For instance, Goldman (2009: 80) also argues that an analogous too fine-grained criterion of individuation of belief-forming methods raises trouble for Williamson's safety condition on knowledge, and so does Whitcomb (2008).<sup>11</sup> In a similar spirit, and more to the point given the cases of interest here, Ranalli (2014: 1231-1234) observes that, if understood with reference to the agent's actual and factive belief-forming method, safety inevitably delivers a knowledge verdict in **Fake Barn**. Here's why. Given factivity, it's logically impossible to see a barn without there not being a barn: accordingly, in all the nearby worlds where the agent sees a barn and believes there is a barn the agent forms a true belief. The safety condition is satisfied and the verdict in **Fake Barn** has to be knowledge. Similar problems arise also in the context of the safety dilemma: as a result, the conciliatory strategy leaves the dilemma unsolved.

---

<sup>9</sup>On this point, terminological clarification is needed. Need a proper functioning accurate clock be as accurate as the veridical perception of the only real barn? Perhaps the clock's hands are slightly off and therefore slightly inaccurate, and yet the clock seems to be still functioning properly. But by 'properly functioning clock' I mean a properly functioning accurate enough clock. A properly functioning clock need not (and often will not) be perfectly accurate: even if the hands are slightly off, it will be accurate enough to read the correct time. This imperfect degree of accuracy will apply to the belief-forming method in **Fake Barn**: properly functioning perceptual capacities need not (and often will not) be perfectly accurate (perhaps they won't discriminate some details of the shape or colour of the barn), but they're nonetheless accurate enough to single out the real barn. Parity of reasoning here requires two very fine-grained and ultimately factive methods: reading from a properly accurate clock and having an accurate perception of the real barn. Thanks to an anonymous reviewer of *Synthese* for encouraging me to be more precise on this.

<sup>10</sup>For more discussion of *de re* barn beliefs, see Bernecker (2020: 1107), Hiller and Neta (2007: 312), Pryor (2004: 71), Hawthorne (2005: 56, fn. 17) and Brown (2000).

<sup>11</sup>See Williamson (2009: 305-308) for a reply to Goldman's criticism.

For sake of clarity and to make the problem more vivid, I'd like to offer an additional case. A genuinely perceptual version of **Frankfurt Clock** brings out even more clearly the difficulties that the conciliatory strategy incurs. Consider the following case:

**Bowl of Grapes.** You and your flatmate are both very hungry, so you go to the kitchen to have some grapes. You desperately want those grapes, but your flatmate has an interest that you don't eat them, so they're willing to do whatever it takes to eat the grapes first and replace them with fake grapes. However, your flatmate is slow, so you get to the kitchen before them. Everything is in order, and you see the bowl of grapes on the table, but had you arrived a few minutes later, you'd have seen the fake grapes.

This is a perceptual variation of **Frankfurt Clock**.<sup>12</sup> To capture the knowledge verdict by appeal to safety, advocates of the conciliatory strategy must specify the belief-forming method in a fine-grained and factive fashion, and then include the real grapes in the relevant method: if the method is specified as merely seeing grapes, the belief turns out false in nearby worlds. But once the same belief-forming method is operating in another perceptual case like **Fake Barn**, then barn beliefs will also be trivially safe, and the two cases of interest are collapsed. In fairness, I should note that the advocate of the conciliatory strategy may still point out another principled difference between the two cases, and then individuate belief-forming methods in a fine-grained, narrow and ultimately factive manner in **Frankfurt Clock** and in a more coarse-grained, non-factive manner in **Fake Barn**. However, absent any such difference and motivation, this individuation of belief-forming methods is clearly *ad hoc*.

Let's take stock. So construed, safety fails to distinguish between the two cases: both Russell's and Barney's beliefs are safe, and thus candidates for knowledge. This is especially troublesome once we appreciate that one of the main motivations for adopting a full-blooded safety condition is precisely to explain absence of knowledge in fake barn cases. If this minimal desideratum is not met, safety loses its appeal. Overall, relativizing safety to more fine-grained belief-forming methods does not make progress on the safety dilemma: a successful solution is to be sought elsewhere.

### 3.4 Environments

Pritchard's approach to Kelp's dilemma is revisionary: he defuses Kelp's challenge by sacrificing the knowledge intuition in **Frankfurt Clock**. Pritchard maintains that such intuition mistakes mere cognitive achievement with knowledge: while Russell forms a true belief on the basis of his cognitive ability (that is, reading a clock), his belief falls short of knowledge courtesy of environmental luck. Pritchard writes:

---

<sup>12</sup>For the case to go through, I am stipulating that experiences can have high level content: seeing that object *o* is a grape involves the concept GRAPE, which is plausibly a rich/high-level fruit concept. See Siegel (2021: Section 4.3 and Section 6) for relevant discussion. Thanks to an anonymous reviewer of *Synthese* for urging me to clarify this point.

What is being exhibited in this example is, I would claim, not knowledge at all, but rather a mere cognitive achievement. *In order to see this, one only has to note that this case is essentially a barn-façade-style case in which the agent's true belief is infected by environmental luck.* While nothing intervenes between [Russell's] cognitive ability and his cognitive success—he really does employ his cognitive abilities in order to gain his true belief about the time, *he is in a very unfriendly environment from an epistemic point of view.* (Pritchard 2009: 40; emphasis mine)

In what follows, I take issue with Pritchard's claim that the agent in **Frankfurt Clock** is in an unfriendly environment from an epistemic point of view. I will argue that there's an important modal difference between *actually* unfriendly environments and only *potentially* unfriendly environments. While the former may be incompatible with knowledge, we have little reason to think that also the latter are. Since it will be crucial also to the refined formulation of safety that I outline in the next section, this modal difference between the two environments is worthy of special attention.

To get a better purchase on Pritchard's diagnosis, let's focus on environmental luck. In **Fake Barn**, (veritic) environmental luck makes Barney's belief only luckily true for reasons having to do with specific actual features of the environment. The belief is true but it could easily have been false because of the environment in which it was formed (more precisely, because of the many papier-mâché copies of the real barn). Elsewhere too, Pritchard assumes that knowledge and environmental luck are incompatible (e.g., Pritchard 2009: 38; Pritchard 2012a: 273).

If **Frankfurt Clock** is also a genuine case of environmental luck, then it surely is a very unusual one. Canonical instances of environmental luck correspond to actually unfriendly environments: the possibility of error is not abstract, but it's rather concretely realised in some specific feature of the environment under consideration. We see no such feature in **Frankfurt Clock**. Since the demon could but ultimately does not intervene, the environment is fine from an epistemic point of view. Granted, the environment could have been unfriendly, the demon could have intervened. However, absent such intervention, the environment is at best potentially unfriendly, not actually unfriendly. There is an overlooked but important modal difference between the environment in **Fake Barn** and the environment in **Frankfurt Clock**: while only the former is actually unfriendly from an epistemic point of view, the latter is not, it only could have been.<sup>13</sup>

This modal difference is starker when we consider other clear instances of environmental luck. Take Kvanvig's Comanche case. While the debate is different, the point to drive home is the same: actually unfriendly environments are incompatible with epistemic states like knowledge or understanding. Let's recall Kvanvig's case and pay special attention to its structure:

---

<sup>13</sup>Kelp (2018: 52) briefly notices this difference, but he doubts that it helps the safety theorist to meet his challenge. As I argue in the next section, this difference has far-reaching consequences for safety-based diagnoses of the cases.

**Comanche.** Suppose you pick up a textbook on Native American History and read through a chapter documenting the Comanche dominance of the southern plains, until eventually you seem genuinely to understand why the Comanches dominated the southern plains. But suppose as well that while the book you happened to pick up is accurate, *most other books on this topic are full of errors*. If you had picked up one of these other books instead (*and we can imagine that they are all within easy reach!*), your beliefs about the Comanches would have been almost entirely false. (Kvanvig 2003: 197, 198; emphasis mine)

Bracketing whether it is convincing, Kvanvig's case of environmental luck also features an actually unfriendly environment (there are many inaccurate books on the shelf) rather than a mere potentially unfriendly environment (there could have been inaccurate books on the shelf). The point is general, and applies to further clear-cut instances of environmental luck. Sosa (2010: 471) offers yet another case in point: "Take a subject who drinks from a cup out of several available on a table. All the other cups, let us say, contain a drug that much degrades 'subitizing' ability." This further instance of environmental luck also features an actually unfriendly environment (the cup actually contains the drug) rather than a mere potentially unfriendly environment (the cup could have contained the drug).

In light of these points, To deny knowledge in virtue of environmental luck, Pritchard needs the following version of Kelp's Frankfurt case:

**Fake Clocks.** At 8:22, Russell comes down the stairs and checks the time by looking at one of the many clocks in the room. While most of them are *actually* stopped, Russell happens to look at the only properly functioning clock. As a result, he forms the true belief that it's 8:22.

I take this to be a more compelling instance of environmental luck. The environment is epistemically unfriendly because of an actual live possibility of error. At the same time, it's also clear that **Fake Clocks** and **Frankfurt Clock** are not structurally similar, and Pritchard needs such structural similarity to vindicate his no-knowledge verdict.

I have argued that the environment in **Frankfurt Clock** is not a genuine instance of environmental luck; now I want to focus on why this kind of environment is after all compatible with knowledge. I have previously dubbed it 'potentially-unfriendly-but-actually-friendly'. Here's an explication to make further sense of it: we can think of the environment in **Frankfurt Clock** as an instance of *higher-order environmental luck*. This is because the agent is merely lucky to not be unlucky: they find themselves in an environment that could have been unfriendly, but actually isn't. We see this difference at work by comparing the standard version of **Fake Barn** with this crucially different variation of the case, which I slightly readapt from Carter (2016: 5):

**Fake Barn\*.** Barney\* is driving in the countryside, and hits a fork in the road. If he goes left, he enters a normal epistemically friendly environment.

If he goes right, he enters Fake barn country. He luckily takes a left, sees a barn and forms the belief that there is a barn.

This is a more explicit case of higher order environmental luck – a type of luck that doesn't undermine knowledge. After all, higher-order environmental luck is just an instance of benign luck. Barney\* may be lucky to be in a friendly environment and thus in a position to know that there is a barn. However, his belief is not true courtesy of luck. Similarly, the environment in **Frankfurt Clock** is also affected by higher-order environmental luck. Russell may be equally lucky to have stepped into a friendly environment and to be in a position to know the time, but this doesn't prevent him from coming to know the time. These considerations tell against Pritchard's assessment: **Fake Barn** and **Frankfurt Clock** display a relevant modal difference in their respective environments. The former environment undermines knowledge, but the latter doesn't.

A quick summary of the cases covered so far will be helpful. First, to treat **Frankfurt Clock** on a par with **Fake Barn**, Pritchard needs a different scenario featuring an actually unfriendly environment. I have offered such scenario in **Fake Clocks**. The upshot is that **Frankfurt Clock** and **Fake Barn** are not analogous. Second, I have argued that **Frankfurt Clock** is more similar to **Fake Barn\*** due to a common instance of higher-order environmental luck, which is compatible with knowledge. Pace Pritchard, if the agent lacks knowledge in **Frankfurt Clock**, it's not because of knowledge-undermining environmental luck. Pritchard's diagnosis is inadequate.

At this point, it's also worth pausing to set the record straight on the dialectical situation. Merely appreciating the inadequacy of Pritchard's diagnosis doesn't quite let the safety theorist off the hook yet. One fundamental question remains: does the difference in the two environments help the safety theorist to capture the verdicts of knowledge and ignorance in **Frankfurt Clock** and **Fake Barn**? Once we incorporate this difference into the general formulation of safety, the answer is 'yes'. In fact, the two vignettes display two crucially different environments. Forming a true belief on the basis of a good method (reading from a clock) in a potentially-unfriendly-but-actually-friendly environment (**Frankfurt Clock**) is not the same as forming a true belief on the basis of a good method (visual perception) in an actually unfriendly environment (**Fake Barn**). In the next section, I reformulate the safety condition in light of this difference and I show how it successfully meets Kelp's challenge.

### 3.5 Environment-relative Safety

This is my proposed modification to the safety condition on knowledge:

**Environment-relative Safety.** In most or all close possible worlds in which S believes that  $p$  via the same method of belief formation  $M$  that S uses in the actual world (**sub-condition M**) and S occupies the same environment  $E$  that S occupies in the actual world (**sub-condition E**),  $p$  is true.

Two points are immediately worth noting.

Firstly, one might have qualms about my use of ‘same’. In response, I offer the following remarks. First, methods and environments are not literally identical, but only relevantly similar (compare: different tokens of the same type are relevantly similar but not literally identical to each other). Second, modal epistemologists notoriously struggle to identify satisfactory criteria of individuation of methods and environments across relevantly close world. Some authors have even suggested that said criteria ultimately rest on a modal fallacy (Hetherington 2013: 166-167). While I’m aware of the problems with the notion of similarity and with how to individuate methods and environments in nearby worlds, I also notice that they affect any modal condition on knowledge and do not bear directly on the formulation of safety on offer here.

Secondly, and crucially, this safety condition is different from environment-relative ability conditions that further entail a safety condition (e.g., Greco 2020a: Section 3). The focus here is exclusively on safety and environments, not on abilities. Accordingly, this version of safety differs from the SSS safety condition outlined in Chapter 2.

After clarifying these two key points, Several ideas in this refined formulation of safety deserve unpacking. In what follows, I will:

- (a) Explain and motivate the conjunctive formulation of Environment-relative Safety.
- (b) Clarify how Environment-relative Safety provides a satisfactory solution to the safety dilemma and also deals with further variations of **fake barn**.
- (c) Show that, despite some important limitations, Environment-relative Safety is preferable to the conciliatory and revisionary strategies discussed above.

### 3.6 Safety as a Conjunctive Condition

To bring out the distinctiveness of my formulation of safety, it might be useful to start by comparing it with other versions of the principle. For instance, Williamson’s safety condition reads as follows: “if one knows, one could not have easily been wrong in a similar case” (Williamson 2000: 147). Pritchard (2007: 281) writes: “S’s belief is safe iff in most nearby possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world the belief continues to be true”. Finally, Sosa (1999: 146): “If S were to believe that  $p$ ,  $p$  would be true”.<sup>14</sup>

All these formulations are rather simple: instead of a conjunction of conditions, they feature only one condition. Second, they are relativized, more or less explicitly, to methods or bases only (e.g., Williamson ?: 128; Pritchard 2008: 446; Sosa 2007: 26). Differently, we can think of Environment-relative Safety as the conjunction of two sub-conditions: just as a conjunction is true if and only if both conjuncts are true, so Environment-relative

---

<sup>14</sup>Here, I am following Ranalli (2014: 230, fn. 12) for a similar (and helpful) list of formulations of the safety principle. See also Blome-Tillman (2020: 36) for an appendix with another list of formulations of safety.

Safety is satisfied if and only if both sub-conditions are met. The first, sub-condition M, is familiar: safety theorists have already noted the importance of keeping the belief-forming method fixed. The second, sub-condition E, is new: it comprises environments, and it has received comparatively less attention. Crucially, sub-condition E expands the set of factors that determine which worlds count as relevantly close. As we have seen, these factors standardly include the subject S, the proposition P, the time T, and the method M: S, P, T, M. In the revised version, we keep also the environment E fixed, and enrich the set accordingly: S, P, T, M, E.

This addition distinguishes our two cases of interest, and, far from being *ad hoc*, it also better captures the spirit of the safety condition on knowledge. Safety is supposed to explicate the sense in which knowledge equals non-accidentally true belief. Such ‘non-accidentality’ has to do with the modal profile of both the method employed and the environment in which it is employed. To fully appreciate this point, consider a non-epistemic case of safety. For example, if we aim to assess whether a car is safe, we will make sure to drive it in a suitably specified environment (say, in appropriate driving conditions). In general, judgements on safety tacitly assume environmental factors, so it’s worth making these factors explicit in the formulation of the safety condition on knowledge. To count as safe, a true belief must meet each of the two sub-conditions: if it meets just one, then it fails to constitute knowledge. This is the key to solve the safety dilemma, as I shall explain next.

### 3.7 A New Solution to the Safety Dilemma

Environment-relative Safety not only accommodates presence of knowledge in **Frankfurt Clock**, but it also captures the ignorance intuition in **Fake Barn**. Let’s take each in turn. Presence of knowledge in **Frankfurt Clock**: if we keep the method (competently reading from a clock) and also the potentially unfriendly but actually friendly environment fixed, then Russell’s belief is safe. In the close worlds where Russell competently reads from a clock and the demon doesn’t intervene on the environment, the belief is true and satisfies Environment-relative Safety. Both sub-conditions E and M are met: knowledge is present.

On to ignorance in **Fake Barn**: if we keep the method (visual perception) and also the actually unfriendly environment fixed, then Barney’s belief is unsafe. In the close worlds where Barney continues to occupy an environment filled with fake barns, the belief is false and fails to satisfy Environment-relative Safety. Sub-condition E isn’t met: the environment is and remains unfriendly, so knowledge is absent.

This point brings out a key difference between this formulation of safety and Sosa’s aptness condition, which also stresses the importance of environmental factors. A belief is apt if and only if it’s accurate because adroit (competent) and for the adroitness to explain the accuracy, the believer’s exercise must succeed in a way that overcomes the influence of certain environmental interferences (for instance, Gettier-style deviant causal chains). However, Sosa’s aptness condition remains compatible with some environmental

interferences: in **Fake Barn**, the agent's belief comes out as apt and constitutes (animal) knowledge (Sosa 2010: 473; 2015: 81). Despite the close possibility of an environmental interference, their belief is accurate (true) because it's adroit (competent). Differently, and unlike Sosa's aptness condition, Environment-relative Safety yields a no-knowledge verdict in **Fake Barn**: precisely in virtue of the close possibility of an environmental interference, the agent's belief comes out as unsafe and thus fails to constitute knowledge. When it comes to environmental interferences, the safety condition defended here is more demanding than Sosa's aptness condition, which at best entails a weaker type of safety relative only to the agent's abilities (see Ranalli 2014: 1232 and Greco 2020a: 5152-5154 for discussion). Crucially, my formulation of safety focusses on environments directly, bypasses abilities entirely and thus differs from Sosa's aptness condition.

We now have a key environmental difference between the two cases, and now safety can detect it. In **Frankfurt Clock**, the error possibility requires a change in the environment: for Russell to form a false belief, the demon has to intervene. In **Fake Barn**, no such change is needed. The error possibility is already realised in the environment: because of the actual papier-mâché copies of the real barn, Barney's belief could too easily be false. So construed, safety rises to Kelp's challenge.

However, this formulation of safety incurs further difficulties when it comes to individuate the relevant environments more precisely. For example, indexing safety to the agent's environment and the agent's actual position in that environment may yield the result that the agent knows in fake-barn cases: if we restrict the safety condition to all the worlds where Barney's environment is exactly as it and where he's located where he's actually located, these are all worlds where Barney is looking at the only real barn. Barney seems to have knowledge on such safety principle, and this is bad news for the safety condition I am offering here.

One moral to draw from this type of case is that, just like belief-forming methods, environments should not be individuated too narrowly. Barney's actual environment is broadly misleading: while it does include one real barn, it is more generally constituted by one real barn amidst many fake barns. If we keep such broadly misleading environment fixed, in these close worlds he ends up forming a false belief because in addition to the real barn there's many more fake barns. A knowledge verdict trades on a too narrow individuation of belief-forming methods (Barney forms his belief on the basis of seeing the real barn) but also on an equally narrow individuation of the relevant environment (Barney's location in the only portion of fake barn country where the real barn is visible). As I've argued already, neither of these approaches is ultimately correct.

It's important to be clear on the limits of this proposed safety condition, and testing it against an even harder case does so in an especially vivid way. Imagine a variation of **Fake Barn** in which Barney travels through different regions: some are genuine barn regions and some are fake barn regions. Does Barney have knowledge in genuine barn region? Does he lack knowledge in fake barn region? And what about borderline cases where he's still in genuine barn region and yet very close to fake barn region? This variation

of **Fake Barn** raises a difficult question: which specific concept of environment can suit Environment-relative Safety? Crucially, providing an informative answer to this question turns out to be very challenging.

In order to make progress on this question, we might initially try to equate environments with the spatio-temporal regions of the actual world where the agent is located and then keep these regions fixed across close possible worlds.<sup>15</sup> With this conception of environments at hand, we could argue as follows: given Environment-relative Safety, Barney has knowledge as long as he's located in an actually friendly environment even if he's close and could easily have ended up in an epistemically unfriendly and misleading environment. But this way of identifying environments is deeply problematic because it fails to capture important intuitions concerning the modal closeness of relevant error possibilities. In fact, when Barney merely gets too physically close to the unfriendly environment it looks like he automatically counts as being in the unfriendly environment even if he hasn't crossed the physical border and he's not yet in the physical region filled with fakes. As such, this concept of environments is ill-suited for the safety condition on offer here. The crucial question pertaining to the relevant concept of environment still remains, but what more can be said in answer to this question? While there seems to be at least two available options, neither is fully adequate. Let's briefly review each option in turn.

The first option takes the relevant concept of environment as primitive: perhaps we have a pre-theoretic, context-sensitive concept of environment that we tacitly appeal to when making judgements about cases of knowledge in various environments. However, since it doesn't say much on the key features of knowledge-conducive environments, this line of response is not very informative, and, accordingly, not very satisfactory.

Building on Williamson's circular version of safety, the second line of response consists in providing a closely related circular characterisation of environments: just like we use intuitions about knowledge to guide determinations of safe belief (Williamson 2009: 305), we use our intuitions about knowledge also to guide determinations of what counts as an environment suitable for knowledge. The circularity involved here is not obviously vicious: it just takes the concept of environment suitable for knowledge as being dependent on the concept of knowledge. Perhaps in some key cases we can give only a somewhat circular answer to the question of whether the agent's environment is suitable for knowledge: this is because we first have to answer the question of whether the agent knows, and then use it to that to say something more informative about the relevant features of their environment.<sup>16</sup> While it still leaves important questions open, this circular characterisation of environments can initially suit the safety condition offered here.

---

<sup>15</sup>This concept of environment draws inspiration from the closely related notion of local reliability of belief-forming methods: roughly, a belief-forming method is more or less locally reliable depending on the specific spatio-temporal regions in which the agent is located while employing such method. See Goldman (1986: 45; 2008: 91), Goldberg (2010: 52), Graham (2014a) and Bernecker (2020: Section 2) for a thorough discussion of the notion of local reliability.

<sup>16</sup>This circular account is not necessarily a concession to the knowledge-first programme: knowledge can still be a type of safe true belief formed in a suitable environment even if we ultimately need to consult the concept of knowledge to say something more informative on the features of such environment.

However, neither the primitive nor the circular option seem fully satisfactory: more work needs to be done to see whether each option can deal with further and potentially more challenging cases. Taken together, these considerations suggest that Environment-relative Safety inevitably runs into some version of the generality problem raised against the individuation of environments instead of belief-forming methods: it is difficult to provide a sufficiently specific and ultimately informative account of environments that suits this version of safety. This variation of the generality problem is certainly pressing, but it doesn't affect the solution to the safety dilemma on offer: even without a fully worked out conception of suitable environments for knowledge, we can still register an important environmental difference between the two cases discussed here and use this difference to address Kelp's challenge. Giving a more precise account of what exact features an environment must display in order to be compatible with knowledge (perceptual or otherwise) is a task best left for another occasion. For now, we can more modestly focus on how Environment-relative Safety fares when compared to the other solutions offered in response to the safety dilemma. I do so in the next section.

### 3.8 Theoretical Virtues

Importantly, Environment-relative Safety does better than the conciliatory and revisionary strategies discussed above. First, unlike the conciliatory solution, Environment-relative Safety does not have to specify the belief-forming method in a too specific fine-grained way. Instead of focussing on a properly functioning clock and collapsing the two cases of interest, the relevant method is simply Russell's general ability to read a clock. Second, unlike the revisionary solution, Environment-relative Safety is not committed to a dubious conception of environmental luck, and it straightforwardly captures the correct intuitions of knowledge and ignorance in each case.

I also wish to flag that Environment-Relative Safety is not the only available solution to Kelp's dilemma. Beddor and Pavese's (2018) and Wedgwood's (2020) normality approaches also provide a principled distinction between the environment in **Frankfurt Clock** (absent any demon intervention, quite normal) and the environment in **Fake Barn** (since it's filled with mere barn façades, quite abnormal). However, since it's not wedded to the admittedly murky concept of normality, Environment-relative Safety is theoretically simpler, and thus preferable. Moreover, as Kelp (Manuscript) points out, these normality accounts still face a slightly different version of the safety dilemma.

In conclusion, a further index to both methods and environments makes great progress on the safety dilemma. While I acknowledge that such additional index incurs substantial difficulties in individuating the relevant environments, I also maintain that it's the most promising strategy to rescue the safety condition from Kelp's troubling conjunction of cases.

### 3.9 Chapter Summary

By conjoining **Frankfurt Clock** with **Fake Barn**, Kelp raises a thorny dilemma for safety-theoretic accounts of knowledge. He draws a pessimistic conclusion: no formulation of safety can be weak enough to grant knowledge in the former and strong enough to predict ignorance in the latter. The shortcomings of the conciliatory and revisionary solutions further confirm Kelp's conclusion: since they fail to capture the right verdict in one of the two cases, it seems that the safety theorist has no choice but to sacrifice one of these intuitively correct verdicts.

Yet, appearances are misleading. There's a crucial difference between a potentially unfriendly environment (**Frankfurt Clock**) and an actually unfriendly environment (**Fake Barn**). This difference requires to index the safety condition not only to belief-forming methods, but also to environments, and Environment-relative Safety does just that. True enough, this new version of safety is not without problems and leaves many important questions open: crucially, Environment-relative Safety fails to individuate environments in a precise manner and it doesn't fully clarify the range of environments in which agents can fall prey to epistemic Frankfurt cases. However, Environment-relative Safety also provides a principled way out of Kelp's dilemma: in fact, it can predict knowledge in **Frankfurt Clock** and ignorance in **Fake Barn**. To do away with the safety condition on knowledge we may have to wait another day.

# Chapter 4

## Against Kairetic Explanationism

### 4.1 Overview of the Chapter

As we have seen in Chapters 1 and 2, recent epistemology has witnessed a substantial opposition between two competing approaches to capturing the notion of non-accidentality in the analysis of knowledge: the explanationist and the modalist. Because explanationism promises to deliver an analysis of knowledge without any appeal to modal notions, it looks like an elegant proposal with *prima facie* appeal. However, appearances are misleading: in this chapter I raise some objections to a recent explanationist analysis of knowledge and defend a novel version of modalism. I conclude with a reassessment of the explanationist's initial ambition to provide an analysis of knowledge without modal notions. The upshot will be that even if the prospects for such analysis remain dim, our money should be on the modalist, not the explanationist.

### 4.2 The Competition: Safety and Kairetic Explanationism

While the task of providing an analysis of knowledge was once crucial to epistemology, in retrospect it looks more like a hopeless series of misplaced efforts. Troubles began when Gettier's paper was taken to show that even if truth, belief and justification are individually necessary, they are nonetheless jointly insufficient conditions for knowledge. This naturally prompted the question at the very heart of the Gettier problem: in addition to these three, which further condition is ultimately sufficient for knowledge? Epistemologists embarked in a long journey searching for such missing condition, and considered different candidates along the way. Modal conditions stood out among the most promising but equally fell prey to Gettier-style counterexamples, and, after almost three decades of unsuccessful proposals, the hope in a full-blooded analysis slowly but steadily wore off.<sup>1</sup>

---

<sup>1</sup>For an overview of post-Gettier epistemology, see Ichikawa and Steup (2018). For a more historically informed perspective, see Antognazza (2015). For discussion of Gettier-style counterexamples raised specifically against the sufficiency of modal conditions on knowledge, see in particular Goldberg (2015: Section 1) and Grundmann (2018: Section 3).

Craig (1990) urged to focus on the function of the concept of knowledge rather than on its analysis, Zagzebski (1994) showed that the Gettier problem is virtually inescapable, and Williamson (2000: 2-5) suggested, quite radically, that knowledge simply cannot be analysed. As a result, many epistemologists no longer hold out hope that there is – or there even can be – a satisfactory analysis of knowledge, and such pessimism appears to be well-motivated.

Many, though not every epistemologist: some don't share the pessimism and still retain hope. Recent times have seen the rise of novel explanationist analyses of knowledge.<sup>2</sup> The explanationist begins by taking issue with a dominant figure of post-Gettier epistemology, the modalist. With Sosa (1999), modalists ask how knowledge must be *modally* related to what is known. They give two answers: sensitivity (roughly, S knows that *p* if and only if in the closest possible world where *p* is false, S does not believe that *p*) or safety (roughly, S knows that *p* if and only if in most or all close possible worlds S's belief that *p* is true).<sup>3</sup> Modalists occasionally quibble over the details, but they all agree that a modal condition like sensitivity or safety is the key to solve the Gettier problem. And at precisely this point, the explanationist demurs: instead of asking how knowledge must be modally related to what is known, she asks how knowledge must be explanatorily related to what is known. Since explanation is a hyperintensional notion (Nolan 2014: 157), the key to solve the Gettier problem is not a modal condition. According to the explanationist, knowledge doesn't require the subject's belief to be true in an appropriate set of possible worlds. More simply, it only requires the subject to believe *p* because *p* is true.

In this chapter, I object to the explanationist and I side with the modalist: I shall argue that the most recent explanationist condition offered by Bogardus and Perrin ([forthcoming](#)) remains insufficient for knowledge due to, once again, Gettier-style cases. But the significance of my contribution goes beyond a mere discussion of the intuitively correct verdicts in Gettier-style vignettes, for what is at stake is more than just that. If successful, this recent explanationist proposal would deliver what many epistemologists have longed for – an analysis of knowledge – and it would do so without modal conditions. Crucially, the success of explanationism also casts serious doubt on the relevance of modal notions to epistemological theorising, and, *a fortiori*, on the tenability of modalism in general. Accordingly, modalists have urgent enough reason to halt the explanationist advance, and such is the task I take on here: I will show that the most recent version of explanationism not only fails to provide a satisfactory analysis, but it also incurs difficulties that modalism easily avoids. Ultimately, such version of explanationism does not live up to its promises: thus, epistemologists should not do away with modal conditions on knowledge – not yet at least.

Here's a more precise overview of the chapter. In section 1, I lay out the main tenets of explanationism and focus on the putative analysis of knowledge offered by its most recent

---

<sup>2</sup>For instance, Bogardus and Perrin ([forthcoming](#)), Faraci (2019), Korman and Locke (2020). Less recent advocates include Alan Goldman (1984), Rieber (1998), Neta (2002), Jenkins (2006).

<sup>3</sup>Nozick (1981), Sainsbury (1997), Sosa (1999, 2007), Williamson (2000), Pritchard (2005, 2007, 2012a), Kelp (2013), Pavese and Beddor (2018), Clarke-Doane and Baras (2021).

advocates (Bogardus and Perrin [forthcoming](#)). In section 2, I argue that the explanationist condition is not sufficient for knowledge due to standard Gettier cases and fake-barn cases. The latter are especially troubling in virtue of a problematic closure failure. In section 3, I turn the spotlight to modalism and I defend it in two stages: I recall the refined safety condition developed in the previous chapter and I show its superiority in handling the same Gettier cases that beset Bogardus and Perrin’s version of explanationism. I close by reassessing the case for their initial ambition to provide an analysis of knowledge without modal notions. The upshot will be that even if the prospects for such analysis remain dim and modalism is not problem-free, our money should be on the modalist, not the explanationist.

### 4.3 Kairetic Explanationism Outlined

Explanationist analyses of knowledge take the following simple form: S knows that  $p$  if and only if S believes that  $p$  because  $p$  is true. The truth of  $p$  must *enter prominently* – or *figure crucially* – in the explanation of why the target belief is held. Instead of modal conditions and related possible world talk, the focus is primarily on the *hyperintensional* notion of explanation.<sup>4</sup> Explanationist approaches may vary in the details, but they all accept some qualified version of this simple analysis.

Variation in the details is due to how the key notion of explanation is unpacked. Alan Goldman (1984: 101) and Rieber (1998) both invoke the truth of  $p$  as “the best explanation” of why S believes  $p$ . However, as also noted elsewhere (Jenkins 2006: 142-143), both these analyses leave the central notion of explanation too vaguely stated. Accordingly, I shall put them aside for the purpose of this paper. Carrie Jenkins (2006: 139) refines the explanationist condition: according to her *scopic* notion of explanation, she appeals to the truth of  $p$  being a good explanation for an *outsider* – someone not acquainted with a more specific set of facts. Despite her explanationist sympathies, Jenkins is nonetheless keen to deny the ambitious status of reductive analysis to her proposal. More modestly, she only aims at “*getting a handle* on the concept of knowledge” and “*understanding what its role in our lives might be*” (Jenkins 2006: 138; emphasis mine). Since my focus is on full-blooded analyses of knowledge and Jenkins is careful to stress that she’s not offering one, I will have to put also Jenkins’s proposal to one side.

This leaves me with one main candidate left. In what follows, the action will be exclusively on the most recent and fully developed explanationist analysis offered by Bogardus and Perrin ([forthcoming](#)). My reason to do so is simple: Bogardus and Perrin ([forthcoming](#): 1) are explicit in stating that their explanationist condition is both necessary and sufficient for knowledge. Hence, theirs is a full-blooded reductive analysis. Since I am

---

<sup>4</sup>Recall our discussion in Chapter 1. Hyperintensional notions distinguish between necessarily equivalent contents and block substitution *salva veritate*. It is widely agreed that explanation is hyperintensional: there’s a truth-conditional difference between “I am in danger because Doctor Jekyll is in the room” and “I am in danger because Mr. Hyde is in the room”, even though Doctor Jekyll and Mr. Hyde are intensionally equivalent terms. For a state-of-the-art overview of hyperintensionality, see Berto and Nolan (2021).

chiefly interested in whether explanationism provides a satisfactory reductive analysis and they are indeed offering one, I shall narrow down my focus specifically to their proposal.<sup>5</sup>

Bogardus and Perrin make a two-fold case for explanationism. First, they offer a knockdown counterexample aimed at every modalist approach. Then, they proceed to provide an alternative explanationist analysis of knowledge. Before focussing on their explanationist analysis, let's look at the knockdown counterexample to modalism. Consider the following case already introduced in Chapter 2:

**Atomic Clock.** The world's most accurate clock hangs in Mia's office. The clock's accuracy is due to a clever radiation sensor. However, this radiation sensor is very sensitive and could easily malfunction if a radioactive isotope were to decay in the vicinity. This morning, against the odds, someone did in fact leave a small amount of a radioactive isotope near the world's most accurate clock in Mia's office. This alien isotope has a relatively short half-life, but – quite improbably – it has not yet decayed at all. It is 8:20 am. The alien isotope will decay at any moment, but it is indeterminate when exactly it will decay. Whenever it does, it will disrupt the clock's sensor, and freeze the clock on the reading "8:22." Therefore, though it is currently functioning properly, the clock's sensor is not safe. The clock is in danger of stopping at any moment, even while it currently continues to be the world's most accurate clock. (Bogardus 2014: 12; Bogardus and Perrin forthcoming: 4)

According to the explanationist, Mia knows the time but her belief is neither safe nor sensitive. Mia's belief is unsafe: since the isotope could have easily decayed, in the very close worlds where it does the clock stops and Mia forms the false belief that it's 8:22. Mia's belief is also insensitive: in the closest possible world where the isotope decays and the clock stops, she forms the false belief that it's 8:22. Thus, had  $p$  been false, Mia would have believed it anyway. From this, the explanationist swiftly concludes that no modal condition is necessary for knowledge.<sup>6</sup> Such conclusion paves the way for a different explanationist analysis, to which I now turn the attention.

We can see this explanationist analysis in action by focussing on ordinary cases of perceptual knowledge (this will be important also for the related explanationist treatment of fake-barn cases; more on this below). Suppose I truly believe that there's a cup of coffee in front of me. There is a set of different explanations for my belief: it might be because it appears to me that there's a cup of coffee, because of particularly favourable

---

<sup>5</sup>A quick but important terminological clarification: unless specified otherwise, I will henceforth use 'explanationism' and 'the explanationist' to refer to Bogardus and Perrin's specific version of kairetic explanationism.

<sup>6</sup>In the explanationist's own words: "Such a case nicely pries apart our concept of knowledge from our concepts of safety and sensitivity, by polluting very many of the 'nearby' worlds with false beliefs while maintaining a tight enough connection between belief and its truth to allow for knowledge" (Bogardus and Perrin forthcoming: 4). For more counterexamples against the necessity of safety for knowledge, see Neta and Rohrbaugh (2004), Comesaña (2005) and Kelp (2009, 2016). I return to this counterexample in section 3, and I show it to be toothless against my version of modalism.

lighting conditions, because of my properly-functioning visual system, because I am particularly craving coffee at that time of the day, and so on. If we start to run the kairetic test, we quickly notice how none of these explanations really makes a difference: we can subtract each of these candidate *explanantes*, and yet, according to Bogardus and Perrin, the explanation wouldn't be complete. For example, if the explanation offered for why I (truly!) believe that there's a cup of coffee in front of me is that it only appears that there is one, something would be missing because in fact there truly is a cup of coffee in front of me. To reach a complete and satisfactory explanation, we must appeal to the truth of *p*: I believe that there's a cup of coffee in front of me because it's true that there's a cup of coffee in front of me. The truth of *p* crucially figures in the explanation of why the target belief is held: hence, the belief under consideration constitutes knowledge. According to the explanationist, the same procedure generalises beyond perceptual knowledge and it's meant to apply, *mutatis mutandis*, to further cases of interest to epistemologists (e.g., knowledge from memory, testimony and inference).

With these points in play, it's now worth pausing to appreciate the ambitions of Bogardus and Perrin's proposal. Equipped only with the notion of explanation and Strevens' kairetic test, they set out to cover all the relevant instances of knowledge: perceptual, inductive, deductive, moral, and mathematical (Bogardus and Perrin [forthcoming](#): Section 3). Moreover, given cases like **Atomic Clock**, they also reject the key modalist assumption, namely any relevance of modal conditions on knowledge. They are quite explicit on this; for sake of vividness, consider the following passage:

We believe modalism misses something crucial about the nature of knowledge: *the connection between a believer and the truth can't be fully captured in modal terms, because it's an explanatory connection*. Modalism, then, can have no long-term success as a research project. It's standing in its own grave. So, it's time to look elsewhere for the analysis of knowledge. (Bogardus and Perrin [forthcoming](#): 7, emphasis mine)

This is the main ambition of Bogardus and Perrin's explanationism: to provide a successful reductive analysis of knowledge without modal notions. Crucially, for their analysis to be successful, the explanationist condition must suffice for knowledge and thus be immune from Gettier-style counterexamples. Yet, as I argue in the next sections, the explanationist analysis has no such immunity and falls prey to Gettier-style cases. Modalism, on the contrary, does not stand in its own grave: it is not undermined by counterexamples like **Atomic Clock** and it easily accommodates the same Gettier cases that will be shown to beset explanationism. Defending the superiority of modalism will occupy section 3; now I want to focus on two familiar vignettes featuring stopped clocks and fake barns.

## 4.4 Kairetic Explanationism Exposed: Stopped Clocks

Consider the following run-of-the-mill Gettier case:

**Stopped Clock.** Russell takes a competent reading from a clock that he knows to be reliable and has no reason to think is currently not working. Based on this reading, Russell forms the belief that it is 8:22 pm. What's more, the belief is true: it is indeed 8:22 pm. There is, however, a twist to the story: the clock is broken and the reason Russell's belief is true is that the clock happened to stop working exactly twenty-four hours ago.<sup>7</sup>

Intuitively, Russell has a justified true belief that it's 8:22, but he doesn't know it. Can the explanationist accommodate this verdict? In this initial case, the answer is 'yes'. Russell believes that it's 8:22 pm because the clock reads 8:22 pm. However, the clock reads 8:22 not because *it is* 8:22. Rather, it reads 8:22 because *it was* 8:22 exactly twenty-four hours before (Bogardus and Perrin [forthcoming](#): 17). Russell doesn't believe that *p* because *p* is true: the bi-conditional of the explanationist analysis is not satisfied, and, accordingly, the explanationist condition is not met. Since the truth of *p* doesn't figure crucially in the explanation of why the target belief is held, the target belief does not constitute knowledge. So far, so good: explanationism can capture absence of knowledge in this familiar Gettier-style case and it secures the right result.

But now consider the following slightly different version of **Stopped Clock**:

**Defective Stopped Clock.** Russell takes a competent reading from a clock that he knows to be reliable and has no reason to think is currently not working. Based on this reading, he forms the belief that it's 8:22 pm. What is more, it is 8:22 pm and the clock correctly reads 8:22 pm. There is, however, a twist to the story: *in virtue of an undetected manufacturing defect, the clock is designed to stop at exactly 8:22 pm, which is also when Russell happens to look at it.* It's 8:22 pm, the clock stops at 8:22 pm, and Russell truly believes that it's 8:22 pm.

Here too, Russell may have a justified true belief that it's 8:22 pm, but he doesn't know it. The reason why Russell fails to know is the uncontroversial assumption that one can't know the time from a stopped clock *regardless* of when exactly the clock happens to stop. Can the explanationist secure the right result also in this case? Unfortunately, now the answer is 'no'. Russell believes that it's 8:22 pm because the clock reads 8:22 pm. Unlike the previous version of this Gettier case, the clock reads 8:22 pm because it is 8:22 pm. Nevertheless, the clock also stops because it is 8:22 pm, and one hardly comes to know the time by consulting a stopped clock. Russell believes that *p* because *p* is true: the truth of *p* crucially figures in the explanation of why the target belief is held. However, the same fact also explains why the clock stops: given the undetected manufacturing defect, the fact that it's actually 8:22 pm is also the reason why the clock stops. Russell believes that *p* because *p* is true, but, *p* the explanationist, he intuitively fails to know.

---

<sup>7</sup>Following Pritchard (2012a), this is a standard Gettier case featuring intervening luck: the agent's belief is initially false, but it then hits upon the truth courtesy of a stroke of luck. Differently, non-standard fake-barn cases display environmental luck: the agent's belief is true, but it could have easily been false given the environment where it was formed.

The explanationist analysis fails to pass the sufficiency test with more refined versions of standard Gettier cases. The explanationist condition is met: Russell believes that it's 8:22 pm because the clock reads 8:22 pm, the clock reads 8:22 pm because it is 8:22 pm and it's true that it is 8:22 pm. Yet, knowledge is intuitively absent. After all, you can't come to know the time from a stopped clock.<sup>8</sup> Even granting that the explanationist condition is necessary, **Defective Stopped Clock** shows that it remains insufficient for knowledge. Standard Gettier cases like this call into question the adequacy of explanationism. As I proceed to show next, non-standard Gettier cases like fake-barn cases do the same and raise additional difficulties for the explanationist condition.

## 4.5 Kairetic Explanationism Exposed: Fake Barns

Recall another Gettier-style case, offered by Alvin Goldman (1976) but credited to Carl Ginet:

**Fake Barn.** Barney, a reliable barn spotter, is driving through the countryside. He looks out of the window, sees a barn on the hill and comes to believe that he is looking at a barn. Whilst Barney's belief is true, unbeknownst to him, the structure he is looking at is the only real barn in an area filled with fake barns that are indistinguishable from real barns.

What can the explanationist say about **Fake Barn**? Bogardus and Perrin (forthcoming: 16) tentatively suggest a no-knowledge verdict.<sup>9</sup> However, their treatment of the case is ambiguous between two different *de re* and *de dicto* readings: once disambiguated, their diagnosis turns out problematic because of an especially troublesome closure failure. Let me take each point in turn and start with the details of their diagnosis of **Fake Barn**.

Bogardus and Perrin begin by considering an ordinary case of seeing a barn, and treat it on a par with equally ordinary cases of perceptual knowledge. In such ordinary cases, this is the relevant explanation: "Barney believes that there's a barn on the hill because

---

<sup>8</sup>Or can you? Explanationists may bite the bullet and hold that the agent knows in **Defective Stopped Clock**. I anticipate the following problems with this response. First, linguistic evidence tells against utterances like "I took a reading from a stopped clock, and I came to know the time". Upon reflection, this knowledge attribution sounds implausible. Second, the case features the double stroke of luck typical of Gettier-style vignettes: one bad (the clock stops) and one good (the clock stops at an only accidentally correct time). If knowledge is absent in the original version, it should be absent in this structurally similar version too. Perhaps the explanationist could treat the case as an unusual instance of testimony: Russell gains knowledge from an unreliable source as in other alleged cases of unreliable testimonial knowledge (Hawthorne 2005: 68; Goldberg 2005). While I take this to be the most promising response, I think it rests on a too inclusive (and hence distorted) conception of testimony: the case doesn't involve a single speech act, arguably a key defining feature of testimonial knowledge. For reasons of space, I will not pursue this issue further.

<sup>9</sup>For the purpose of this chapter, I shall grant to Bogardus and Perrin that some no-knowledge verdict in **Fake Barn** is correct. However, I also note that both epistemologists (Lycan 2006; Sosa 2010: 472-473; Turri 2016; Schellenberg 2018: 211) and experimental philosophers (Colaco et alia 2014) have adduced arguments to challenge the correctness of the no-knowledge verdict in **Fake Barn**. Accordingly, I care to emphasise that, unlike **Defective Stopped Clock**, the ensuing objections do not clearly apply to other versions of explanationism that provide a different diagnosis of fake-barn cases. I am grateful to an anonymous reviewer of *Episteme* for pushing me to clarify this point.

it looks like there's a barn on the hill and it's true that there's a barn on the hill". The truth of  $p$  (that there is a barn on the hill) crucially figures in a complete explanation of why Barney believes that there is a barn on the hill; accordingly, in this ordinary case, Barney knows. However, in **Fake Barn**, a different explanation enters the picture. It goes as follows: "Barney believes that there's a barn on the hill because it looks like there's a barn on the hill. However, it looks like there's a barn on the hill not because it's true that there's a barn on the hill but only because he is in an environment where most objects look like barns". Otherwise put, Barney believes that there's a barn on the hill because he finds himself in Fake Barn Country. To back up such potentially surprising diagnosis, Bogardus and Perrin offer the following remarks, worth quoting in full:

Even if your eyes happen to fall upon a real barn in a forest of fakes, we might begin to think it false that it looks like there's a barn before you because there is a barn before you. As the barn facades proliferate, a rival explanation looms into view: *that it looks like there's a barn before you because you're in a region full of structures that look like barns*. In other words, it becomes plausible to say that you believe there's a barn before you because it looks like there's a barn before you, and it looks like there's a barn before you *because you're in Fake Barn Country*. (Bogardus and Perrin 2014: 16; emphasis mine)

This is how Bogardus and Perrin aim to capture the no-knowledge verdict in **Fake Barn**: unlike ordinary instances of perceptual knowledge, in this case the truth of  $p$  doesn't figure crucially in the explanation of why the agent believes that there's a barn. The agent doesn't believe that  $p$  because  $p$  is true: rather, they believe that  $p$  because  $p$  *appears* to be true. According to Bogardus and Perrin, in **Fake Barn** we ought to explain Barney's true belief by appealing to the fact that objects look like barns, not the fact that they are real barns. Hence, no knowledge.<sup>10</sup>

It's worth noting that Bogardus and Perrin's explanationist diagnosis seems *prima facie* hard to make sense of. Let's zero in on the alleged "rival explanation" of the belief under consideration: why exactly should we explain Barney's true belief by appealing to the fact that "he's in a region full of structures that look like barns"? The explanation sounds *ad hoc*, and it doesn't go far enough. Suppose that Barney is back in real barn country, where barns look like barns and are actual barns. Should we also explain Barney's true belief by appeal to the fact that he's in a region full of structures that look like barns? If the explanation doesn't stop at appearances in the good case of real barns, it's unclear why it should stop at appearances in the 'bad' case of seeing the only real barn in the mists of papier-mâché copies. After all, in both cases the agent believes that there's a barn because it looks like there's a barn and it's true that there's a barn. The two appearance-based explanations are equally incomplete: Bogardus and Perrin's explanationism seems

---

<sup>10</sup>Another early advocate of explanationism, Alan Goldman (1984: 44), also agrees: "In such cases ... the proper explanation for the belief appeals to the broader context of the perceiver's being in the vicinity of all these look-alike objects, any of which would produce the belief in question." Bogardus and Perrin acknowledge Alan Goldman's point in a footnote (Bogardus and Perrin [forthcoming](#): 16, footnote 23).

to lack the resources to provide a principled difference between the two cases. The main motivation for the no-knowledge verdict starts to lose plausibility.

But even granting that their explanationist diagnosis is correct (as I shall charitably grant), it applies at best to one version the case. For sake of clarity, it is worth distinguishing between two versions of **Fake Barn** – *de re* and *de dicto* respectively.<sup>11</sup>

In the *de re* version, Barney comes to believe that that object on the hill is a barn. Barney's demonstrative belief is distinctively about *that* object, that particular barn on the hill. In the *de dicto* version, Barney comes to believe, more simply, that there is *a* barn on the hill. The indefinite article is crucial. Barney's *de dicto* belief is not demonstrative: it is not about that particular barn, it is just the belief that there is a barn on the hill.

With this distinction in play, let's revisit Bogardus and Perrin's explanationist diagnosis. Consider the *de re* version of **Fake barn**. In this case, the explanationist has to grant knowledge to *de re* barn beliefs: in fact, they are structurally identical to ordinary cases of knowledgeable perceptual beliefs. Barney truly believes that that very object on the hill is a barn because that very object looks like a barn and it's true that that very object is a barn: Barney forms a demonstrative true belief because that very object looks like and is a barn, not because other objects look like barns. Thus, the explanationist is committed to presence of knowledge *de re*, just like they predict presence of knowledge in the good case of seeing a barn and in general instances of perceptual knowledge. In general, explanationists do not have much of a choice here: since it would carry over to ordinary perceptual cases, a no-knowledge verdict in the *de re* version of **Fake Barn** would open the door to a significant (and hence disturbing) sceptical threat.

Now, on to the *de dicto* version. In this case, Bogardus and Perrin's explanationist diagnosis may apply: Barney truly believes that there's a barn on the hill not because it's true that there's a barn on the hill, but because many other objects in that portion of the environment look like barns. If the explanationist aims to capture absence of knowledge in **Fake Barn**, then it inevitably has to be absence of knowledge *de dicto*. Thus, to sum up, Barney knows *de re* but fails to know *de dicto*.

This is, however, a problematic result. Suppose that some version of the closure principle on knowledge is true. For instance,

**Closure.** If one knows P and competently deduces Q from P, thereby coming to believe Q while retaining one's knowledge that P, then one comes to know that Q. (Williamson 2000, Hawthorne 2005)

The principle enjoys independent plausibility, but defending it exceeds the scope of this paper. For my purposes, I shall more simply point out that if we conjoin the *de re* and *de dicto* versions of **Fake Barn** then this independently plausible principle fails. Let's fill in the details as follows: let P and Q be the *de re* and *de dicto* the barn propositions respectively. Barney knows (*de re*) that that very object on the hill at is a barn, he competently

---

<sup>11</sup>For more discussion of *de re* barn beliefs, see Bernecker (2020: 1107), Hiller and Neta (2007: 312), Pryor (2004: 71), Hawthorne (2005: 56, fn. 17) and Brown (2000).

deduces that since that object is a barn then there's a barn on the hill but, according to Bogardus and Perrin, he fails to know (*de dicto*) that there's a barn on the hill. Borrowing DeRose's (1995) terminology, their explanationist analysis is committed to the following 'abominable' conjunction: Barney knows that that object on the hill is a barn but he does not know that there's a barn on the hill. We shouldn't accept this result light-heartedly: how could Barney know that that object on the hill is a barn and not know that there's a barn on the hill?<sup>12</sup>

Let me make the same point, but from a different angle. Suppose further that some version of the knowledge norm of assertion is correct. For example,

**KNA.** One must: assert *p* if and only if one knows *p*.

This norm has prominent advocates (e.g., Unger 1975, Williamson 2000, Sosa 2015, Kelp and Simion 2019), but defending it here would also take me too far afield. For my purposes, I shall more simply note that according to Bogardus and Perrin's version of explanationism Barney could permissibly assert the following: "I know that that object on the hill is a barn, but I don't know that there's a barn on the hill". Given Barney's knowledge *de re*, this sounds infelicitous. Moreover, if Barney asserted the *de dicto* proposition that "There's a barn on the hill", according to Bogardus and Perrin's explanationist condition he would strictly speaking violate the knowledge norm of assertion (though perhaps only blamelessly). This is also implausible: Barney seems intuitively entitled to make this assertion, even more so if we concede with Bogardus and Perrin that he knows *de re* that that object is a barn. Yet another result we should not accept light-heartedly.<sup>13</sup>

On these last two points, I want to tread carefully. My objection does not rest on accepting the closure principle or the knowledge norm of assertion: the problem has to do with the too simple explanationist diagnosis of very hard fake-barn cases. Even granting that Barney fails to know in **Fake Barn**, this applies only to the *de dicto* and not to the *de re* version of the vignette. Once we conjoin these two distinct versions, Bogardus and Perrin's explanationist diagnosis turns out untenable. It may well be that the closure principle sometimes fails, but this does not seem an acceptable case of closure failure. It may well be also that the infelicity of Barney's assertion can be explained away by further pragmatic considerations, or that there is no theoretical pressure to square an analysis of knowledge with an independent linguistic norm of assertion. However, granting knowledge *de re* and denying knowledge *de dicto* remains problematic, and this is because Bogardus and Perrin's explanationist analysis doesn't provide a complete and plausible treatment of the **Fake Barn** vignette. The reason may be this: fake-barn cases are notoriously difficult, and put pressure on very complex analyses of knowledge. McGlynn (2014: 173) nicely makes this point as follows:

Barn cases seem to bring out just how *demanding* a state knowledge really is;

---

<sup>12</sup>For similar reasons, Bogardus and Perrin's version of explanationism also struggles with Kripke's red barn example (Kripke 2011: Chapter 7).

<sup>13</sup>See Hawthorne (2005: 32) for a defence of closure principles of knowledge based on considerations of conversational propriety.

the moral seems to be that knowing that  $p$  makes demands on one's external environment to a greater degree than we might have otherwise expected.

His remarks equally apply to the explanationist analysis: knowing that  $p$  makes demands to one's external environment to a greater degree than just believing  $p$  because  $p$  is true. Accordingly, it should not be so surprising that Bogardus and Perrin's explanationist condition cannot properly deal with them. On this front too, their version of explanationism fails to deliver.

Let me take stock. In the last two sections, I have argued that Bogardus and Perrin's explanationist condition remains insufficient for knowledge. In **Defective Stopped Clock**, their condition is met but knowledge is absent. In **Fake Barn**, the no-knowledge verdict applies only to the *de dicto* version of the case, and this consequently leads to a problematic closure failure. Overall, the explanationist analysis fails to provide a satisfactory diagnosis of the two main types of Gettier-style cases.<sup>14</sup> As I proceed to argue next, the different safety approach outlined in Chapter 3 handles the very same cases rather easily.

## 4.6 Environment-relative Safety Defended

For my defence of modalism, I propose and endorse this formulation of the safety condition (slightly modified from the previous chapter):

**Environment-relative Safety.** In most or all close possible worlds in which  $S$  believes that  $p$  via the same method of belief formation  $M$  that  $S$  uses in the actual world and  $S$  occupies the same environment  $E$  that  $S$  occupies in the actual world,  $p$  is true

Two points of clarifications are in order. First, like other formulations of safety, this version allows for safety failure in case of false beliefs in different propositions than  $p$ . The focus is not on whether a belief-forming method  $M$  employed in an environment  $E$  yields true beliefs only in  $p$ , but also in propositions relevantly similar to  $p$ .<sup>15</sup> Second, and unlike other formulations of safety, this version is relative to both methods and environments. Such double index better captures the spirit of the safety condition on knowledge. Safety theorists aim to assess whether a specific enough belief-forming method would produce

---

<sup>14</sup>To forestall misunderstanding and for sake of charity, I shall briefly consider a line of response on behalf of the explanationist. In a short footnote, Bogardus and Perrin ([forthcoming](#): 17, footnote 27) claim that since warrant entails truth, under explanationism there are no justified beliefs without knowledge and thus, a fortiori, no Gettier cases. This move does not let the explanationist off the hook. Firstly, factive justification incurs what Kelp (2018: 84) helpfully dubs the 'new' Gettier problem, the problem to explain why agents lack justification in Gettier-style cases. On this point, explanationists remain suspiciously silent. Secondly, and more importantly, even granting that Gettiered agents lack justification, they nevertheless fail to know: at best, they have a true belief rather than a justified true belief. The explanationist condition remains insufficient for knowledge regardless of whether justification is present. Overall, this move does not help the explanationist with the cases considered in this paper.

<sup>15</sup>Brown (2000), Williamson (2000) and Pritchard (2012a) advocate similar versions of global-method safety. See Rabinowitz (2011), Bernecker (2020: Section 3) and Hirvelä (2019) for discussion.

true beliefs in relevantly similar situations: to make such assessment, it's crucial to focus on the method and the environment where the method is employed.<sup>16</sup> Thus, let's ask: were S to form beliefs via the same method M in the same environment E, would she continue to believe truly? With the proper focus on the relevant worlds, the answer is 'yes'. In **Atomic Clock**, the error-possibility requires a change in the environment: the isotope has to decay. If we keep both method and environment fixed and we focus on the worlds where the subject reads from the atomic clock and the isotope doesn't decay, then the subject continues forming true beliefs. The safety condition is satisfied: accordingly, the knowledge verdict is aptly captured. Let's make the point with reference to a graphic aid (viz. figure 4.1 and figure 4.2 below):

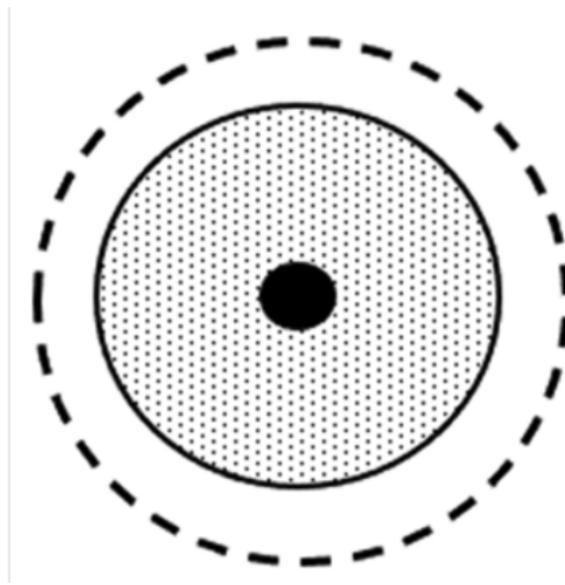


Figure 4.1: The dotted sphere represents the close worlds where only the method is held fixed. In these worlds, the belief under consideration is false and unsafe.

Following Lewis (1973; 1986), let's imagine possible worlds falling into concentric circles centred on the actual world (the black sphere), with "proximity serving as a [visual] metaphor for similarity" (2016 2016: 106). Figure 4.1 represents the **Atomic** counterexample. The wide red sphere includes the worlds where only the method (reading from the atomic clock) is held fixed: since the isotope decays, in these worlds the subject's belief is false. Figure 4.2 highlights the benefits of a double index to methods and environments. In the smaller green sphere, we see the close similar worlds where both the method (reading from the clock) and the environment (the isotope does not decay) are held fixed: in such worlds, the subject's belief is true. Crucially, the belief turns out false only in the further red sphere, namely in the less similar worlds where the method is fixed but the

<sup>16</sup>A word of caution on the methodology adopted here. I invoke subjunctive conditionals because they give us a better grip on which worlds count as relevantly close. By attending to these properly formulated conditionals and focussing on their truth-values, we get a better (though not perfect) understanding of the modal profile of the belief under consideration. Crucially, this is not the only way to flesh out the safety condition: rather, it's just a useful heuristic to achieve clearer judgements about especially controversial cases. See Bogardus (2014: 6-9) for similar remarks and Smith (2016: 111-116) for an illuminating discussion on the modal structure of the safety condition on knowledge.

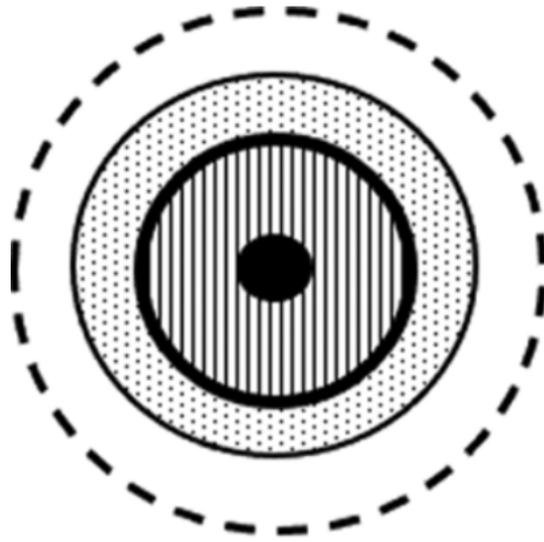


Figure 4.2: The striped sphere represents the close worlds where both the method and the environment are held fixed. In these worlds, the belief under consideration is true and safe.

environment changes. In the similar close worlds in the green sphere, the subject's belief remains true: for the belief to be false, we need to travel out to the worlds in the next sphere, which may be similar but not similar enough to threaten a properly understood safety condition.

In **Atomic Clock**, the explanationist focusses on worlds that may seem intuitively close but nonetheless require a change in the environment: this change places them further away from the actual world. Thus, the case is not strong enough to motivate a full rejection of modalism. It's easy to come up with error-possibilities, but way harder to evaluate precisely how close they are: given a suitable index to both methods and environments, the error-possibility that the explanationist is appealing to is not after all close enough to render the target belief unsafe. Accordingly, the knockdown counterexample to modalism does not apply to the formulation of safety I am offering here. To do away with modal conditions on knowledge, the explanationist must do better.

Now, let's look at the Gettier cases that explanationism struggled with. Equipped with the refined safety condition canvassed above, the modalist accommodates the right verdict in each vignette. Consider again **Defective Stopped Clock**. Since the clock *actually stops* at 8:22 pm, we should focus on the close worlds where the clock remains stopped. Russell's belief is not safe, and hence does not amount to knowledge: in the close worlds where Russell looks at the stopped clock, he too easily forms a false belief about the time. In fact, Russell too easily forms false beliefs in distinct but relevantly similar propositions. Suppose that, two minutes later, Russell forms the belief that two minutes have passed since 8:22 pm. Since the clock would still read 8:22 pm, Russell's belief in this distinct but relevantly similar proposition is false. Unlike the explanationist analysis, this refined formulation of safety correctly predicts that Russell fails to know.

Next, on to **Fake Barn**. A properly understood safety condition delivers a no-knowledge

verdict in both the *de re* and *de dicto* versions of the vignette. This is due, once again, to the fact that safety is best understood as globalised to a set of relevantly similar propositions. The agent lacks knowledge *de re* because they would easily form a false belief in distinct and yet relevantly similar *de dicto* and *de re* propositions. Barney fails to know that the very object is looking at is a barn because in close possible worlds he could have easily formed the false belief that there's a barn on the hill, or that another barn-looking object is three-dimensional, inhabited by farmers, meant to stock hay, and so on (cf. Bernecker 2020: 5107). Crucially, by denying knowledge to the *de re* version of fake barn, this version of safety avoids the closure failure that the explanationist is inevitably committed to. On this count too, modalism is superior – and hence preferable – to explanationism.

Before concluding, one final and very important caveat. I do not pretend to have shown that the superiority of modalism is *definitive*. I am well aware that no formulation of safety is problem-free: modal conditions on knowledge incur serious (and perhaps unavoidable) difficulties. Modalists face four uphill battles in the form of the following questions. First, in what sense are beliefs, methods and environments relevantly similar? The vagueness of the notion of similarity is a well-known source of problems for modalism. Second, how do we individuate the set of propositions that the safety condition is globalised to? Modalists usually rely on an intuitive grasp of the set of relevantly similar propositions without being too specific on this point. Third, how fine-grained the individuation of the belief-forming method and environments has to be? This is a version of the generality problem, which has plagued modalist analyses of knowledge for a long time. While certainly pressing, I hasten to flag that none of these issues affect my modest defence of modalism: the action here is primarily on Gettier-style cases rather than on the more general problems that affect modalism.

Crucially, as we have seen in Chapter 2, one key aspect where explanationism does generally better than virtually any version of modalism is the case of knowledge of necessary truths which, as such, are true in every nearby world and thus trivially safe. Worse still for the sufficiency of modalist conditions on knowledge, the absence of an explanatory connection between the method employed and the truth of the belief in question suggests that beliefs in necessary truths may be only luckily true. Conversely, given the key focus on explanatory – rather than modal – connections between belief and truth, explanationists easily deal with such long-standing problem afflicting modalist analyses of knowledge.<sup>17</sup>

For all these reasons, a full-fledged defence of modalism exceeds the scope of the chapter and will have to wait until the next one. My aim here is far more modest: all I argue is that when it comes to the Gettier-style cases considered here, my modalist condition easily delivers the intuitively correct verdicts, while Bogardus and Perrin's explanationist analysis clearly does not. This gives us defeasible but strong enough reason to prefer my modalist approach to Bogardus and Perrin's explanationist alternative.

---

<sup>17</sup>See especially Faraci (2019: 12-15), Korman and Locke (2020: Section 3) for discussion of the advantages of explanationism over modalism in the case of beliefs in necessary truths. See Hirvelä (2019) for a promising modalist attempt to deal with the issue.

## 4.7 Chapter Summary

I have considered two competing approaches to the task of capturing the notion of non-accidentality in the analysis of knowledge: the explanationist and the modalist. I focused mostly on the most recent version of the former, and found it wanting. Bogardus and Perrin's explanationist approach promises to offer a full-blooded reductive analysis of knowledge without modal conditions. However, when tested against a broader range of Gettier-style cases, such approach does not live up to its promises. Their explanationist condition is not sufficient for knowledge: in **Defective Stopped Clock**, the agent believes that  $p$  because  $p$  is true, and yet they lack knowledge. Moreover, their version of explanationism fails to provide a satisfactory diagnosis of fake-barn cases: by delivering a knowledge verdict *de re* and an ignorance verdict *de dicto*, Bogardus and Perrin are committed to a troublesome closure failure. I take this to be a more general indication of the inadequacy of the explanationist approach: however unpacked, the notion of explanation does not seem to be well suited to deliver a satisfactory analysis of knowledge. It is of course possible that other versions of explanationism may do better than Bogardus and Perrin's, but in light of the present discussion their specific explanationist proposal remains problematic.

What about the modalist? Far from being perfect, the safety condition canvassed in the last chapters has at least two important virtues. First, it escapes the knockdown counterexample raised against any modalist approach: **Atomic Clock** poses no threat to such refined formulation of safety. Second, it has a relatively easy time in dealing with the same Gettier cases that beset explanationism: the formulation delivers the correct verdict of ignorance in each of the Gettier-style vignettes considered in this paper. Does this suggest that some version of the safety condition ultimately yields a successful analysis of knowledge? I very much doubt it, and epistemologists should not get their hopes up. However, if there exists such an analysis at all, it is more likely to rest on a modal rather than an explanationist condition – or at least so I've argued here.

# Chapter 5

## Modal Proper Functionalism Defended

### 5.1 Chapter Overview

As we have seen in the previous chapters, despite its once substantial appeal the safety condition on knowledge seems to be losing significance. Given a range of powerful counterexamples and the availability of better developed virtue-theoretic, normic and explanationist approaches, these days fewer and fewer epistemologists are willing to accept the necessity or the sufficiency of safety for knowledge. In this chapter, I aim to reverse this trend: by combining it with proper functionalism, I propose a novel formulation of safety and defend it against the most pressing objections raised in the literature. My conclusion will be the following: once safety is suitably understood as a type of modal proper functionalism, it stands a better chance as the right anti-Gettier condition on knowledge.

### 5.2 A New Hybrid

Since the publication of Gettier (1963), epistemologists have widely agreed that even if belief, truth and justification may be individually necessary, they are nonetheless jointly insufficient conditions for knowledge. While this point should be fairly uncontroversial, what still remains controversial is how to best answer to the crucial question raised by Gettier's paper: in addition to belief, truth and justification, which fourth 'anti-Gettier' condition can be necessary and sufficient for knowledge? Many prominent anti-Gettier conditions have been offered, and yet, to this day, such crucial question remains unanswered.

In this chapter, my aim is to revive an old answer to this question – an answer which has now fallen out of fashion. I propose a novel formulation of safety and offer it as necessary and sufficient anti-Gettier condition. I do so by developing a hitherto unexplored hybrid: I shall elucidate safety by appeal to proper function. In keeping with safety-theoretic approaches, the account of knowledge proposed in this paper is modal in that it views knowledge as a matter of modal robustness across relevant possible worlds; however, modal robustness is in turn cashed out in terms of proper function of belief-forming

methods in appropriate cognitive environments. This novel hybrid comes with important advantages. First, in virtue of being indexed to both belief-forming methods and environments, it advances our understanding of the safety condition – so far largely underdeveloped and too uninformative to be satisfactory. Second, not only can it withstand the most pressing objections moved against standard versions of safety, but it also passes the necessity and sufficiency tests for knowledge. My plan is as follows. Section 1 gives a brief survey of the present state of play: I map out the extant versions of safety and I situate them in the context of other prominent modal and non-modal anti-Gettier conditions. Section 2 explains why, given the way it has been usually understood, safety can no longer be a serious contender as necessary and sufficient anti-Gettier condition for knowledge. I also identify the key desiderata that a suitably understood safety condition ought to meet. In Section 3 I put forth and motivate my novel formulation of safety in terms of proper function and I show that it meets these key desiderata. Next, in Section 4, I check my proposed formulation against the most pressing counterexamples to both the necessity and the sufficiency of safety for knowledge, and I detail how it withstands each of them. Section 5 brings out the distinctiveness of my proposed formulation by comparing it with other prominent modal and non-modal anti-Gettier conditions. I conclude with a methodological afterthought, explaining to what extent the safety condition offered in this paper can yield a definitive and successful analysis of knowledge.

### 5.3 The Rise and Fall of Safety

Even if once very popular, the safety condition on knowledge has now lost much of its original significance: in fact, if it features at all in any of the most prominent contemporary accounts of knowledge, it never really takes centre stage. Because the literature on safety is extensive, at first pass we can introduce it by appeal to the following oversimplified but intuitive gloss: a belief is safe if and only if, given the way it was formed, it couldn't easily have been false.<sup>1</sup> Crucially, the 'couldn't' is unpacked *modally* – that is, in terms of relevantly similar close possible worlds. Early advocates of safety include Williamson (1994), Sainsbury (1997), Sosa (1999) and Pritchard (2005). Each of these authors has defended a different version of safety with different motivations, but two general points are worth emphasizing. First, unlike the other popular modal condition on knowledge (sensitivity), safety complies with closure principles and permits to know the denial of sceptical hypotheses. Secondly, safety also explains absence of knowledge in several key cases routinely discussed in epistemology – most importantly Gettier cases and lottery cases.<sup>2</sup>

---

<sup>1</sup>See Rabinowitz (2011) for an overview of the safety condition of knowledge.

<sup>2</sup>Williamson (1994) firstly appeals to a safety-style argument in the context of the problem of vagueness, and then returns to the safety condition for knowledge in later work (Williamson 2000, 2009b). Sainsbury (1997) follows up on Williamson's early work and expands on the notion of easy possibility. Sosa (1999) focusses on a counterfactual formulation of safety which is meant to improve on both Nozick's sensitivity condition (Nozick 1981) and Goldman relevant alternatives theory of knowledge (Goldman 1976). Sosa's counterfactual formulation is not without problems: see Comesaña (2007) and McGlynn (2012) for criticism; see Holliday (2015) for a formal discussion of counterfactual theories of knowledge. Pritchard (2005,

But despite these *prima facie* advantages, at present not many authors would defend a robust version of safety: that is, not many authors would accept that safety is a necessary (let alone sufficient) condition for knowledge. In what follows, I provide a quick overview of the current state of play with a special focus on the role of safety in the most prominent recent accounts of knowledge. I will begin with the accounts that are most sympathetic to safety and end with those that reject it altogether: doing so will make it especially vivid that safety is an endangered epistemological species, and it will also pave the way for the fix that I offer in the later sections.

At the most sympathetic end of the spectrum, we find what is sometimes called ‘impure virtue epistemology’ (Pritchard 2012a, 2016, 2020; Kelp 2013). Impure virtue epistemologists accept some version of the safety condition, but then they supplement it with a *separate* ability condition. While safety does contribute to knowledge, it does so only *in conjunction* with a different ability condition.<sup>3</sup> Clearly enough, according to impure virtue epistemologists, safety can’t stand on its own: hence, it needs to be paired with a distinct ability condition.

On to pure or ‘robust’ virtue epistemologists (Sosa 2007, 2015, 2021; Greco 2010, 2020a; Carter 2016; Turri 2016). Here, depending on the robust virtue epistemologist in question, safety is either *supplanted* or *absorbed* by a more fine-grained ability condition. Sosa no longer accepts a safety condition on knowledge (Sosa 2007), and now opts for an ability condition elucidated in terms of the key concept of aptness (Sosa 2015).<sup>4</sup> Greco (2020a 2010) develops an environment-relative conception of abilities that at best entail some redundant (and weaker) version of safety. Carter (2016) also proposes a more graded (as opposed to rigid) notion of abilities to capture the central intuitions standardly accommodated by the safety condition.<sup>5</sup> But crucially, safety by itself is not doing any of the relevant explanatory work: in fact, it doesn’t even figure in these accounts of knowl-

---

2007) develops a safety condition in order to deal with cases of veritic luck – the type of luck which makes a belief only accidentally true and prevents it from constituting knowledge. Pritchard’s safety condition also purports to explain absence of knowledge in Gettier and lottery cases. As should be clear, each of these authors develops the safety condition for different reasons; some of these authors have also revised their view during the years (e.g., Sosa and Pritchard). For present purposes, it’s important to note the diversity of the main formulations of safety and the key reasons that make it attractive.

<sup>3</sup>Impure virtue epistemologists pair safety with an ability condition for reasons having to do with counterexamples to the sufficiency of safety for knowledge (e.g., TrueTemp; see Pritchard 2012a). I return to these cases later on in the chapter and I show how to deal with them without supplementing safety with a separate ability condition.

<sup>4</sup>This is not wholly uncontroversial: Greco (2020a) disagrees and argues that Sosa is in fact consistent when it comes to reconcile the apparent tension between safety and aptness.

<sup>5</sup>Robust virtue epistemologists notoriously struggle with cases of environmental luck exemplified by Ginet’s/Goldman’s famous fake-barn scenarios. While the agents in these scenarios do seem to successfully exercise their cognitive abilities, they nevertheless intuitively lack knowledge because of the misleading nature of the environment (Goldman 1976; Pritchard 2012a). Robust virtue epistemologists have either bitten the bullet and conceded knowledge in fake-barn cases (Sosa 2015) or appealed to a more fine-grained conception of abilities to capture the no-knowledge intuition (Carter 2016; Littlejohn 2014). I also note that fake-barn cases are especially controversial: the no knowledge intuition has been challenged on both philosophical (Lycan 2006, Schellenberg 2018) and experimental grounds (Colaço et alia 2014). I return to these issues when I discuss what my proposed safety condition predicts about these cases; for now, it’s important to note that even if an ability condition entails some version of safety, such safety condition will be too weak to explain absence of knowledge in fake-barn cases, and the key explanatory work will be done by some distinct ability condition. Yet again, safety is an idle wheel.

edge. Indeed, according to pure robust virtue epistemologists, knowledge is primarily an achievement which may or may not entail safety: for this reason, in their accounts the role of the safety condition proves ultimately marginal.

A significantly different and thus stand-alone version of virtue epistemology is Pavese and Beddor's modal virtue epistemology (Pavese and Beddor 2018). While they are sympathetic to safety, just like robust virtue epistemologists Pavese and Beddor also hold that some version of safety is merely entailed by their ability condition restricted to normal worlds (Pavese and Beddor 2018: 68). As a result, also this version of virtue epistemology does away with safety and focusses on a more explanatory central ability condition. Once again, safety is merely a redundant – and therefore less significant – condition on knowledge.

Since I mentioned normal worlds, it's worth moving on to the recent new wave of normalcy epistemologies (Peet and Pitcovski 2018, Goodman and Salow 2018, Littlejohn and Dutant 2020, Nado and Horvath 2021).<sup>6</sup> While most of these views retain the spirit of modal conditions on knowledge, the explanatory focus is on the key notion of normality, which is in turn unpacked differently depending on the specific version of normalcy epistemology. It is therefore fair to say that, despite the general agreement with the safety theorist that knowledge is essentially a modal matter, in these normalcy accounts the safety condition is simply replaced by a different normalcy condition. In these accounts too, safety is absent.

We have now reached the opposite end of the spectrum – views that are unsympathetic to safety and reject it altogether. According to more or less recent explanationist approaches, it is the (non-modal) notion of explanation that holds the key to deliver a successful analysis of knowledge (Jenkins 2006, Neta 2002, Bogardus and Perrin forthcoming, Bernecker forthcoming). Explanationists are well aware that since their central notion of explanation is hyperintensional (Nolan 2014: 157), it follows that knowledge “cannot be captured in purely modal terms” (Bogardus and Perrin forthcoming: 7). Advocates of explanationism have adduced a variety of motivations for their approach.<sup>7</sup> But given their discontent with modal conditions on knowledge in general, they also reject *a fortiori* any relevance of safety for knowledge.

Before ending this section, I want to briefly focus on what we may call the ‘circular’ accounts of safety. These accounts are prominently defended by champions of the knowledge-first approach to epistemology (Williamson 2000, 2009b; Lasonen-Aarnio 2010; Kelp 2017). While these authors accept the necessity (Williamson 2000) or the sufficiency of safety for knowledge (Williamson 2009; Lasonen-Aarnio 2010), they elucidate safety in terms of knowledge rather than the other way around (Williamson 2009b;

---

<sup>6</sup>My focus is on knowledge, but see Smith's extensive work on normic support as a view of propositional justification for a detailed discussion of a modal conception of normality (Smith 2016, 2010).

<sup>7</sup>Some explanationists build on debunking arguments in metaethics (Korman and Locke 2020). Others point out the shortcomings of modal accounts of defeat: for example, Korman and Lock (forthcoming) and Bogardus and Perrin (forthcoming) criticise the modal security account of defeat defended by Clarke-Doane and Baras (2021). Others emphasise the difficulties incurred by modal accounts of knowledge in dealing with the problem of necessary truths (Faraci 2019).

Kelp 2017). As I explained above, I am primarily interested in a safety condition robust enough to serve as successful anti-Gettier condition and make predictions on whether a true belief constitutes knowledge: since these authors are not interested in developing safety as an anti-Gettier condition, safety does not take centre stage in their knowledge-first epistemological accounts.

Let's take stock. As made clear by this quick overview, safety is losing significance: most prominent accounts of knowledge reject safety as an explicit necessary or sufficient condition for knowledge and if it features at all in the remaining accounts, safety plays at best a secondary and ultimately redundant role. Given this rather negative picture, it is natural to ask why exactly the safety condition fell out of fashion. In the next section, I take up this question: I provide a new diagnosis of the shortcomings of extant versions of safety and outline the key desiderata that a better safety condition ought to meet. With an eye on such desiderata, I will then be able to develop and offer such better safety condition in the remainder of the paper.

## 5.4 Diagnosis and Desiderata

Despite the advantages and the motivations that once made safety attractive, we can identify at least two main sources of discontent that substantially contributed to its decline in popularity.

Firstly, there's a serious worry about *extensional adequacy*: the safety condition faces very strong counterexamples to its necessity and sufficiency for knowledge. To cast doubt on the necessity of safety for knowledge, epistemologists have put forth several cases of unsafe knowledge. In these cases, agents intuitively seem to know in the actual world despite forming a false belief in relevantly similar close possible worlds (Neta and Rohrbaugh 2004; Comesaña 2005; Baumann 2008; Kelp 2009, 2016; Bogardus 2014).<sup>8</sup> To put pressure on the sufficiency of safety for knowledge, epistemologists have also offered cases where agents form a trivially safe true belief and yet they lack knowledge: for instance, new versions of meta-incoherence cases à la Mr. TrueTemp (BonJour 1980, Lehrer 1990, Pritchard 2012a) and the long-standing problem with trivially safe necessary truths (Roland and Cogburn 2011). In addition to these counterexamples, other authors have questioned the success of safety as anti-Gettier condition. *De Re* versions of Goldman's fake-barn cases (Pryor 2004), the conjunction of fake barn cases with epistemic Frankfurt cases (Kelp 2016) and more refined Gettier-style cases (Miracchi 2015) strongly suggest that safety fails as satisfactory anti-Gettier condition. All these counterexamples led epistemologists to either implement safety with a separate condition or to abandon it entirely.

So much for extensional adequacy. But on the top of this first worry, there's a second – and perhaps deeper – worry: the safety condition is too vague and uninformative. In

---

<sup>8</sup>See also Sosa (2010: 471) for the contention that the nearness of a dream scenario puts pressure on the necessity of safety for knowledge.

the previous section, safety was glossed as follows: a belief is safe if and only if, given the way it was formed, it couldn't easily have been false (in most or all relevantly close worlds where the subject continues to form beliefs in the same/sufficiently similar way as in the actual world, the subject's belief continues to be true).<sup>9</sup> While this gloss captures both the core idea that modal robustness is essential to knowledge and that safety is best understood as relative to the way the belief is formed (bases, or belief-forming method), this basic formulation also leaves open far too many questions. Here's a few: how do we exactly individuate the relevant possible worlds? What are the appropriate belief-forming methods? Is safety relativized to belief forming methods only, or, just like some conceptions of abilities, should it be relative also to environments? Some advocates of safety such as Williamson (2009) and Pritchard (2012a) hold that safe belief-forming methods yield true beliefs not in one single proposition  $p$  but in a range of propositions sufficiently similar to  $p$ .<sup>10</sup> If so, how can we precisely demarcate the relevant range of propositions? The success of safety crucially depends on these questions, and yet safety theorists have either answered too vaguely or not answered at all.<sup>11</sup>

Taken together, these worries highlight the two main shortcomings of the safety condition. Firstly, it faces too many counterexamples to its necessity and sufficiency for knowledge: it's not extensionally adequate. Secondly, and relatedly, it is also too vague and underspecified: overall, it's not informative enough to serve as a satisfactory necessary and sufficient anti-Gettier condition. This diagnosis suggests two main desiderata for a better safety condition, which ought to be:

- *Extensionally adequate*, and withstand the most pressing counterexamples to its necessity/sufficiency for knowledge and its tenability as anti-Gettier condition.
- *More informative*, and provide a set of exhaustive criteria to individuate (i) the relevant possible worlds, (ii) the appropriate belief-forming methods, (iii) the appropriate environments and (iv) range of propositions.

It's worth pausing to set the record straight on the main upshot of the present discussion. The safety condition remains too underdeveloped: it just offers a useful *template* with

---

<sup>9</sup>This gloss captures the spirit of the most prominent formulations of safety. For instance, Williamson's safety condition reads as follows: "if one knows, one could not have easily been wrong in a similar case" (Williamson 2000: 147). Pritchard (2007: 281) writes: "S's belief is safe iff in most nearby possible worlds in which S continues to form her belief about the target proposition in the same way as in the actual world the belief continues to be true". Finally, Sosa (1999: 146): "If S were to believe that  $p$ ,  $p$  would be true". See Ranalli (2014: 230, fn.12) and Blome-Tillman (2020: Appendix) for a nice taxonomy of formulations of safety.

<sup>10</sup>For discussion on how to identify the relevant range of propositions, see Hirvelä (2019) and Bernecker (2020).

<sup>11</sup>See Pritchard (2008: 444-446; 2013: 158) and Rabinowitz (2011: Section 3) for a brief discussion of individuation of belief-forming methods; see also the exchange between Goldman and Williamson in Williamson (2009). See also Alfano (2009), Grundmann (2018) and Hirvelä (2019). With the exception of Hirvelä's virtue-theoretic criterion, these discussions have focused on fine-grained/coarse-grained or internal/external individuation of belief-forming methods; what is absent is a precise criterion that specifies belief-forming methods by appeal to more than some relevant features of the actual world. I provide such criterion in the later section.

different parameters that must be filled in a more informative and precise manner.<sup>12</sup> Such parameters include (but are not limited to): the proposition(s) believed, the relevantly similar close worlds, the appropriate belief-forming methods and perhaps also the appropriate environments where said belief-forming methods can be employed. More crudely put: to stand a chance as an extensionally adequate, informative and successful anti-Gettier condition which is necessary and sufficient for knowledge, safety must be made more precise. Importantly, a hitherto unexplored way to make the safety condition more precise is to appeal to proper functionalism: such framework provides the necessary explanatory resources to fill all the parameters that the safety template leaves blank. In the next section, I rely on the proper functionalist framework and I proceed to fill such parameters. The resulting safety condition will meet the two desiderata outlined above, and thus stand a much better chance as extensionally adequate, informative and successful anti-Gettier condition.

## 5.5 Modal Proper Functionalism Outlined

Given that safety stands in need of further clarification, it is natural to look for the right explanatory framework that can achieve such vital clarificatory goal. My suggestion is to look at proper functionalism. I do so for two main reasons. Firstly, proper functionalism has a proven track record of success in elucidating key epistemic concepts like *knowledge* (Millikan 1984; Plantinga 1993), *entitlement* (Graham 2012), and *justification* (Bergmann 2006; Simion 2019). Secondly, as I explain in this section, the safety condition is compatible with proper functionalism and perhaps even naturally lends itself to a proper functionalist reading.

Let's begin with the central notion of proper function. This notion is meant to explicate not only the essentially teleological sense in which objects like organisms and artifacts have a certain *function* (they are *for* something; they have a *purpose* or they display a *design*), but also how said objects are *supposed* to function properly in suitably specified circumstances (appropriate environments). For example, a human heart is supposed to pump blood by beating at approximately 70 beats per minute in a healthy enough human body. When it does so, it's functioning *properly* and it fulfils its *purpose*. But how does an item acquire a purpose? When does it function properly, and how is an environment appropriate? According to the influential etiological theory of functions, an item's purpose is equated with the selected effect that explains why the item was replicated through biological reproduction or otherwise.<sup>13</sup> When the item does what it was selected for, it's

---

<sup>12</sup>For sake of vividness, take a moment to compare safety with clearly more complex virtue-theoretic ability conditions (e.g., Sosa's triple A performance normativity framework and triple S account of abilities, or Greco's environment-relative/contextualist account). There's a clear asymmetry here: the vagueness of the safety condition contrasts unfavorably with these more developed and informative virtue-theoretic ability conditions.

<sup>13</sup>While I favour the naturalistic etiological theory of functions defended by Wright (1973), Millikan (1984), Buller (1998), Graham (2012) and Simion (2019), the proposed safety condition can be developed by appeal to different theories of function (e.g., Plantinga's theistic conception; see Plantinga 1993 for a

functioning *normally* and *properly*. And to function normally and properly, the item needs to be located in a suitable normal environment, similar enough to the one where the item originally came to have its function. This normal environment is *appropriate* for the item's function. Setting several complications aside, in order to develop a new version of safety we can focus on these core concepts: purpose as etiological function, proper (normal) function and appropriate (normal) environments.<sup>14</sup>

By appeal to these core concepts, we can fill more precisely and systematically the parameters left empty by the too vague safety template. Recall that a properly understood safety condition ought to be informative and specify all the following: a criterion of individuation for belief-forming methods, the environments where the belief-forming methods are employed, the relevant swath of possible worlds and also the relevant range of propositions. To fill all these parameters, let's plug in the notion of proper function and the etiological theory of functions:

- *Belief-forming methods (bases)*. Properly (normally) functioning belief-forming methods individuated with reference to their etiological function.
- *Environments*. Appropriate cognitive environments for the proper function of the belief-forming methods.
- *Modal Robustness*. All the relevant possible worlds where the proper functionalist conditions for both belief-forming methods and environments are met.
- *Range of propositions*. All the propositions specified in accordance with the etiological function of the belief-forming methods.

Some comments on each of these parameters are in order. As far as concerns belief-forming methods, three points are worth noting. Firstly, the method under consideration ought to be functioning properly. For example, safe perceptual beliefs will result from properly functioning perceptual capacities, and in the case of inferential beliefs formed on the basis of an instrument, said instrument also ought to be functioning properly (it should not display any type of malfunction). Secondly, the belief-forming method must be designed to aim at truth in line with the etiological theory of functions.<sup>15</sup> To wit, a

---

locus classicus). See Graham (2011; 2012: 476, fn. 5) for a detailed comparison of these two theories.

<sup>14</sup>A note on terminology. To avoid confusion with the normalcy views described above, I shall use 'proper function' as 'normal function' and 'appropriate environment' as 'normal environment'. After all, for the etiological theory of functions proper function just is normal function. Graham (2011: 75) also identifies the two. Moreover, despite the mention of the concept of normality, the explanatory work is done by the notions of proper function and the etiological theory of functions. This marks an important difference with normalcy epistemologies.

<sup>15</sup>On the issue of design, I want to tread carefully. For a satisfactory individuation of properly functioning belief-forming methods, all that matters is only that they are designed to aim at truth. What is the exact source of the design is an interesting question, but it doesn't affect the proper functionalist criterion of individuation of belief-forming methods offered here. That said, there are various proposals in the literature, and they all seem *prima facie* compatible with a proper functionalist criterion of individuation of belief-forming methods. Plantinga (1993: Chapter 11) favours a theistic conception of design and opts for conscious intentional design. Following the influential work in philosophy of biology of Wright (1973), Millikan (1984) and Buller (1998), Peter Graham (2011) offers a different naturalistic evolutionary conception of design, and

belief-forming method like reading from a horoscope won't count as such: its etiological function has nothing to do with truth. The same applies also to only accidentally reliable belief-forming methods: for example, forming beliefs on the basis of a fully accurate brain lesion will not count as a properly functioning belief-forming method because, despite the accuracy, such brain lesion fundamentally lacks any function (etiological or otherwise). Thirdly, the etiological function of belief-forming methods plausibly includes a wider range of propositions instead of one single proposition. This applies to both non-inferential belief-forming methods constituted by biological kinds (perceptual systems) and inferential belief-forming methods based on artifacts (instruments). For example, a clock's etiological function is to reliably yield true beliefs in a range of propositions about the current time; a barometer does just the same with a range of propositions about the air pressure. *Mutatis mutandis*, similar considerations also hold for biological kinds: eyesight's etiological function is to reliably yield true beliefs in a set of propositions about the features of one's surroundings. Once again, it is the etiological function of the belief-forming method under consideration that determines the relevant range of propositions.

On to environments. Importantly, proper function of belief-forming methods is not sufficient to elucidate the safety condition. A belief-forming method may be functioning normally, properly, and even perfectly, but if the environment is not appropriate the belief-forming method in question will not yield neither true nor safe beliefs. Beliefs formed on the basis of properly functioning perceptual capacities employed in a broadly misleading environment will not be accurate, nor safe. Accordingly, the safety condition must be indexed not only to properly functioning belief-forming methods, but also to the appropriate cognitive environment for the proper function of such belief-forming methods. Given the etiological theory of functions, these will be suitably normal environments for the normal functioning of the belief-forming method.

Finally, modal robustness. In keeping with safety-theoretic accounts of knowledge, the safety condition developed here requires that the belief under consideration is true in a set of relevant possible worlds. However, modal robustness is cashed out in terms of proper functioning: the relevant possible worlds are those where both the (properly/normally functioning) belief-forming method and the (appropriate/normal) cognitive environment of the actual world are kept fixed. Small changes in the belief-forming methods or environments will make the possible worlds under consideration irrelevant for the proper functionalist safety condition developed here. Putting this all together, we get:

**Safety as Modal Proper Functionalism.** S's belief that  $p$  is safe if and only if in all the possible worlds in which S continues to form her belief in the target

---

adopts an etiological theory of functions according to which functions are equated with biological benefit historically selected across a large enough period of time. Simion (2016b) also offers an etiological theory of functions, but, unlike Graham, she relaxes the historical condition and demands only a more broadly epistemic (rather than specifically biological) type of benefit. As I said, my preferences lie with Graham and Simion's accounts. However, I hasten to flag that questions about the source of the design are orthogonal to the more pressing question for the safety condition – namely how to individuate belief-forming methods. All that is needed for the individuation of belief-forming method is some design and the resulting proper function, regardless of the ultimate source of each.

propositions  $p$  via the same properly functioning belief-forming method  $M$  employed in the actual world (sub-condition  $M$ ) and in the same appropriate cognitive environment  $E$  occupied in the actual world (sub-condition  $E$ ), the belief continues to be true.

Unlike other formulations, this version of the safety condition incorporates the central insights of proper functionalism and the etiological theory of functions. As such, it is indexed to *both* properly (normally) functioning belief-forming methods and appropriate (normal) cognitive environments. Notice finally how this is a *conjunctive* formulation: just like in order for a conjunction to be true *both* conjuncts need to be true, in the same way both sub-conditions  $M$  and  $E$  have to obtain for this safety condition to be met. This reflects the observation made above about the insufficiency of proper function of belief-forming methods: what must be added and specified is the proper function of belief-forming methods in cognitive environments that are appropriate (normal) for the method under consideration.

I want to conclude this section by offering some remarks in support of a proper functionalist reading of the safety condition. I shall do so by highlighting a few relevant conceptual connections between safety and proper function within and outwith epistemology. These connections clarify the hitherto unappreciated compatibility between these two approaches, and also bring out the distinctiveness of the novel formulation of safety I am offering.

Firstly, starting with safety, it's important to note that safety theorists have excellent reasons to appeal to proper function in order to make their modal condition on knowledge more precise. Recall the key idea behind safety: given the way they are formed, safe beliefs are meant to be modally robust, so that they continue to be true across a relevant swath of possible worlds. Crucially, once we understand 'the way they are formed' by reference to properly functioning belief-forming methods employed in appropriate cognitive environments, such beliefs are indeed modally robust: proper function offers the type of (modal) protection from error which is essential to safety. Yet, despite the clear sense in which safety and proper function are compatible and even complementary, safety theorists have never developed their proposed condition by specifying belief-forming methods and environments in terms of proper function. As a result, the safety condition incurred the major difficulties pointed out above.

Secondly, as far as proper functionalism is concerned, there's no clear tension between safety and proper function, or between safety and the etiological theory of functions. Nothing prevents advocates of proper functionalism from incorporating a modal reading of their core notion of proper function.<sup>16</sup> The two approaches are not incompatible. Rather, they

---

<sup>16</sup>The chancy nature of natural selection that contributes to determining the etiological function of belief-forming methods might seem in tension with the anti-luck requirement on knowledge that motivates modal epistemologies. But even granting that evolution is driven by a type of luck, such luck will ultimately be of a benign kind. Borrowing from Pritchard's taxonomy (2005), we can think of the chancy elements of natural selection as a type of *capacity* luck: it is indeed a matter of good luck that belief-forming methods acquired their etiological function, but this doesn't prevent these belief-forming methods from successfully

are complementary and this hitherto unexplored hybrid formulation can be very beneficial to the safety theorists. Finally, and zooming out from epistemology, the combination of safety and proper functionalism appears to be also pre-theoretically plausible because it gains additional support from non-epistemic considerations. Interestingly, judgements about safety seem to align with judgements about proper function. Take for instance the case of a car that we deem to be safe. As such, the car is likely to be functioning properly and to be well designed; conversely, a malfunctioning or poorly designed car will be rather unsafe. Much like the safety condition on offer here, all these judgments also assume some implicit relativization to environments: a perfectly functioning and exceptionally designed car will not be safe if driven on a very icy road – environments always play a key role. Overall, it looks like there are rather compelling reasons to combine safety and proper functionalism both within and outwith epistemology.

Let's take stock. After arguing that the safety condition stands in crucial need of clarification, I've proceeded to make it more precise by appeal to proper function and the etiological theory of functions. For the resulting safety condition, properly functioning belief-forming methods individuated by the etiological theory of functions and employed in appropriate cognitive environments will yield safe beliefs – beliefs that continue to be true in the relevant possible worlds where the standards of proper function of methods and environments are suitably met. So understood, the safety condition offered here is more informative than standard versions of safety: as such, this formulation meets the first desideratum outlined in Section 3. With this novel safety condition at hand, I now proceed to defuse the most pressing counterexamples moved to the necessity and sufficiency of safety for knowledge and the main objections against the success of safety as anti-Gettier condition. Doing so will show how this novel formulation also meets the remaining desideratum: in fact, it can also serve as satisfactory necessary and sufficient condition for knowledge.

## **5.6 Modal Proper Functionalism in Action**

The main objections against safety's extensional adequacy are best sorted into three categories. In the first category we find objections to the success of safety as anti-Gettier condition. Next, we find further counterexamples against its necessity (second category) and sufficiency (third category) for knowledge. I now take each category of objections in turn; I shall start with checking my new safety condition against the two most discussed Gettier-style cases.

### **5.6.1 Gettier Cases: Clocks and Barns**

Here's two familiar Gettier cases, featuring stopped clocks and fake barns:

---

excluding key instances of veritic luck – the type of luck which is incompatible with knowledge. I offer these remarks in support of the compatibility between safety, proper functionalism and the etiological theory of functions.

**Stopped Clock.** You take a competent reading from a clock that you know to be reliable and have no reason to think is currently not working. Based on this reading you form a belief that it is noon. What's more, your belief is true: it is indeed noon. Crucially, however, the clock is broken and the reason your belief is true is that *it happened to stop working exactly twelve hours ago*. (Kelp 2018)

**Fake Barns.** Barney, a reliable barn spotter, is driving through the countryside. He looks out of the window, sees a barn and comes to believe that he is looking at a barn. Whilst Barney's belief is true, unbeknownst to him, *the structure he is looking at is the only real barn in an area filled with fake barns that are indistinguishable from real barns*. (Goldman 1976)

According to version of safety defended here, the beliefs under consideration may be justified and true, but they don't constitute knowledge because, as emphasized in the text, they fail to meet the key proper functionalist conditions on belief-forming methods and environments in the actual world. Accordingly, the beliefs under consideration lack the modal robustness necessary for knowledge. Take for instance **Stopped Clock**. Because the clock is broken, the belief is not based on a properly functioning instrument. As such, the belief under consideration doesn't continue to be true in the relevant possible worlds where the belief-forming method meets a proper functionalist condition. What's more, the belief under consideration could very easily be false: if we keep the (malfunctioning) belief-forming method fixed, the subject forms a false belief in nearby worlds. The lack of proper function explains why the belief under consideration could easily be false: the instrument is malfunctioning, hence the possibility of error. Standard formulations of safety capture absence of knowledge in Gettier cases by emphasizing that the belief under consideration could easily be false, but they don't explain why exactly it could easily be false.<sup>17</sup> By appeal to proper function, we are now better positioned to offer such explanation: in Gettier-style cases agents could easily form a false belief *because of* a cognitive malfunction in their belief-forming method, some inappropriate feature of their cognitive environment, or both.

What about **Fake Barns**? The safety condition defended here predicts absence of knowledge in this much debated Gettier-style case, but instead of relying on some vaguely described intuitions it does so for a principled reason: the no-knowledge verdict is explained by the further index of safety to appropriate (normal) environments. Even if the agent's perceptual capacities are functioning properly, the environment is clearly inappropriate. Since it is filled with mere replicas of real barns, it is too different from the suitable normal environment in which the belief-forming method under consideration (perception) acquired its etiological function. Given the misleading environment where the belief under

---

<sup>17</sup>Here's Pritchard (2013: 156): "Safety can also deal with Gettier-style cases, for these are characteristically cases in which the agent forms (on the same basis as in the actual world) a false belief in the target proposition in a close possible world." But Pritchard does not say more: we are never told why exactly the agent forms a false belief in the target proposition in a close world.

consideration is formed, it could easily be false: if we keep the inappropriate and abnormal environment fixed, the subject forms a false belief in nearby worlds. The modal proximity of error is again explained by appeal to the further index of safety to appropriate normal environments: the belief-forming method is functioning properly, but the environment is not appropriate for the belief-forming method.<sup>18</sup>

This proper functionalist safety condition predicts absence of knowledge in the two most discussed Gettier-style cases (**Stopped Clock** and **Fake Barn**). Each of these cases features a cognitive malfunction or an inappropriate cognitive environment: this impacts on the modal profile of the beliefs under consideration, and it explains why they could easily be false or why they lack the relevant modal robustness which is necessary for knowledge.

## 5.6.2 Necessity

But does a true belief even need to be safe to constitute knowledge? Some epistemologists think that the answer is ‘no’, and have provided several cases of unsafe knowledge. Kelp (2009; 2016; 2018) has pursued this line of argument most consistently. Leaning on Frankfurt’s famous case (Frankfurt 1969), he begins by raising this type of epistemic Frankfurt case in isolation:

**Frankfurt Clock.** Russell’s arch-nemesis, a powerful demon, has an interest that Russell forms a belief that it is 8:22 by looking at the grandfather clock in the hallway when he comes down the stairs. Russell’s arch-nemesis is prepared to do whatever it may take in order to ensure that Russell acquires a belief that it is 8:22 by looking at the grandfather clock when he comes down the stairs. However, Russell’s arch-nemesis is also lazy. He will act only if Russell does not come down the stairs at 8:22 of his own accord. Suppose, as it so happens, Russell does come down the stairs at 8:22. Russell’s arch-nemesis remains inactive. Russell forms a belief that it is 8:22. It is 8:22. The grandfather clock is working reliably as always. (Kelp 2016: 28)

The case is supposed to show that Russell knows the correct time even if he forms a false belief in nearby worlds – he could after all have checked the clock a few minutes later and looked at a stopped clock. Then, Kelp proceeds to pair this case with the already mentioned **Fake Barns**: once these two cases are raised in conjunction, it looks like the safety condition will be either too weak to predict absence of knowledge in **Fake Barns** or too strong to accommodate the knowledge intuition in **Frankfurt Clock**. Borrowing from Kelp’s useful terminology (Kelp 2016), this is the ‘safety dilemma’. Safety doesn’t deliver the right verdict, and given the first horn of the dilemma, it doesn’t even seem to be necessary for knowledge. To back up the first horn of Kelp’s dilemma, we can list more cases of unsafe knowledge. For instance, consider the following similar vignette:

---

<sup>18</sup>ice that neither the belief-forming method nor the environment is so specific as to include the real only real barn: if it did so, safety would be trivially satisfied and thereby fail to capture the no-knowledge verdict in **Fake Barns**. See Pryor (2004), Ranalli (2014) and Bernecker (2020) for discussion.

**Atomic Clock.** The world's most accurate clock hangs in Mia's office. The clock's accuracy is due to a clever radiation sensor. However, this radiation sensor is very sensitive and could easily malfunction if a radioactive isotope were to decay in the vicinity. This morning, against the odds, someone did in fact leave a small amount of a radioactive isotope near the world's most accurate clock in Mia's office. This alien isotope has a relatively short half-life, but – quite improbably – it has not yet decayed at all. It is 8:20 am. The alien isotope will decay at any moment, but it is indeterminate when exactly it will decay. Whenever it does, it will disrupt the clock's sensor, and freeze the clock on the reading "8:22." Therefore, though it is currently functioning properly, the clock's sensor is not safe. The clock is in danger of stopping at any moment, even while it currently continues to be the world's most accurate clock. (Bogardus 2014: 12; Bogardus and Perrin forthcoming: 4)

We notice a common feature: much like **Frankfurt Clock**, the agent could very easily form a false belief – indeed, they almost *do* form a false belief but at the very last moment they don't. Taken together, these two cases suggest that safety isn't necessary for knowledge: even if agents form false beliefs in relevantly close worlds, they nevertheless know. This is bad news for safety in general and, *a fortiori*, also for the safety condition on offer here.

For these cases to go through, it needs to be showed that the beliefs under consideration lack the relevant modal robustness necessary for knowledge. Recall that once safety is elucidated in terms of proper function, all the relevant possible worlds that matter for modal robustness are only those where both the properly functioning belief-forming method and the appropriate environments are kept fixed. For the possibility of error to obtain, all these vignettes assume a change in the environment that then impacts on the proper function of the belief-forming method. In **Frankfurt Clock**, the demon has to *actually* intervene to manipulate the clock. Similarly, in **Atomic Clock**, the isotope has to *actually* decay. However, both the belief-forming method and the environment meet the proper function condition in the actual world, and once both belief-forming methods and environments are kept fixed, the subjects do continue to form true beliefs (after all, they are employing a properly functioning belief-forming method in an appropriate environment). The beliefs under consideration display the modal robustness that is necessary for knowledge according to **Modal Proper Functionalism**: the error possibilities trade on a change in the environment and the belief-forming methods, and such changes place the worlds where the agents form false beliefs further away. As a result, the allegedly close worlds where the agents form false beliefs are ultimately irrelevant, and, since they fully meet the proper function condition, the beliefs under consideration display the modal robustness which is necessary for knowledge. Accordingly, these counterexamples against the necessity of safety for knowledge do not land against the safety condition on offer here.<sup>19</sup>

---

<sup>19</sup>This strategy also applies to Neta and Rohrbaugh (2004), Comesaña (2005) and Baumann 2008 further alleged cases of unsafe knowledge. Notice finally that the safety condition on offer here predicts knowledge

### 5.6.3 Sufficiency

Even granting that safety is necessary, can it also be sufficient for knowledge? There are two main influential objections that suggest otherwise. First, variations of meta-incoherence cases show that trivially safe true beliefs do not amount to knowledge: even if the belief under consideration is true in many close possible worlds, subjects may still not know. Secondly, the problem of trivially safe necessary truths also shows that even if the belief under consideration is true in virtually every possible world, it still doesn't amount to knowledge.<sup>20</sup> Here are these two types of cases:

**TrueTemp.** Temp forms beliefs about the temperature by consulting a broken thermometer. However, Temp has a guardian angel in the room who controls the thermostat, ensuring that the room's temperature matches the reading displayed on the thermometer. Assuming the angel manipulates the thermostat in all nearby worlds, any belief Temp forms about the temperature in such worlds will be true. (Pritchard 2012a)

**Necessary truths.** Emma is using a calculator to count the sum of  $12 \times 13$ . As a result, she forms a true belief that  $12 \times 13 = 156$ . However, the calculator is actually broken and it's generating answers at random. Emma's belief is true in (every) close possible worlds. (Roland and Cogburn 2011; Pritchard 2012a)

Safety theorists have responded to these cases as follows. To deal with **TrueTemp**, Pritchard (2012a, 2016, 2020) has incorporated a distinct ability condition in his analysis of knowledge. Such condition explains why Temp doesn't know (after all, his belief is not true in virtue of the exercise of a relevant cognitive ability). As far as concerns **Necessary Truths**, safety theorists have globalised their safety condition. For a belief-forming method to count as safe, it has to yield true beliefs in a range of sufficiently similar propositions. Accordingly, while Emma's broken calculator yields true belief in one mathematical proposition, given that it's broken, it fails to do so with *other* sufficiently similar but distinct mathematical propositions. However, neither strategy is particularly satisfactory. On the one hand, incorporating a distinct ability condition concedes too much to virtue epistemology, and indirectly suggests that safety fails as stand-alone modal condition on knowledge. On the other, advocates of globalized versions of safety face the difficult task to demarcate with precision the range of sufficiently similar propositions. Moreover, some globalized versions of safety may be too strong and thus not necessary for knowledge.<sup>21</sup> Taken to-

in **Frankfurt Clock** and ignorance in **Fake Barns**; accordingly, it also rises to Kelp's dilemma outlined at the beginning of the section.

<sup>20</sup>Even if they both threaten the sufficiency of safety for knowledge, these cases are structurally different from the Gettier-style vignettes considered above: the justification component essential to Gettier cases is less stark, and the focus is on the triviality of safety (beliefs that are trivially true in most/every possible world) rather than on the type of veritic luck at work in Gettier cases (beliefs that are only *accidentally* true).

<sup>21</sup>Hirvelä (2019) proposes a promising virtue-theoretic criterion of demarcation. Sosa (2004: 180) and Bernecker (2012: 140; 2020: 5108) independently argue that globalized versions of safety are not necessary for knowledge.

gether, these cases are particularly troubling, and have independently contributed to the overall decrease in popularity of safety. As a result, epistemologists have either resorted to distinct conditions (e.g., Pritchard's impure virtue epistemology) or abandoned modal conditions on knowledge altogether (e.g., explanationist epistemologies and their claimed advantage in solving the problem of necessary truths).

What does the safety condition on offer here predict about these troubling cases? Fortunately, a proper functionalist reading of safety correctly predicts absence of knowledge in each. To see why, notice that the vignettes feature clear instances of *malfunctions*: the thermometer and the calculator are broken. Moreover, in at least one case, the environment isn't appropriate for the belief-forming method: thermometers are not supposed to function in environments featuring interventions of guardian angels. Recall that, according to the etiological theory of functions, an item's appropriate (normal) environment must be sufficiently similar to the one where the item came to acquire its function. This crucial environmental condition is clearly not met given the obviously abnormal and inappropriate features of Temp's environment.

With these points in play, here's how the proposed proper functionalist version of safety captures absence of knowledge in each counterexample. Start with **Necessary Truths**. Since the calculator is broken, it fails to fulfil its etiological function. As a result, it also fails to yield true beliefs in the range of propositions that it was designed to yield true beliefs in. Plausibly, a calculator's etiological function is to reliably yield true beliefs in a range of basic mathematical propositions: it will not have extraordinary computational powers, but, when functioning properly, it will be able to do many simple mathematical calculations. This key condition is not met: the broken calculator will yield false beliefs in a range of propositions specified with reference to its etiological function. As such, the belief under consideration is not safe and doesn't amount to knowledge.<sup>22</sup>

Let's move on to **TrueTemp**. As noticed already, the belief under consideration is formed in clearly inappropriate circumstances: thermometers acquire their etiological function in normal environments that do not feature interventions of guardian angels. As a result, Temp's belief lacks the relevant modal robustness: in the relevant possible worlds where the environment is appropriate for the thermometer's etiological function (that is, in absence of guardian angels), Temp's belief is false. Because it lacks the relevant modal robustness which is essential to the safety condition offered here, Temp's belief is not safe and as such it doesn't constitute knowledge.<sup>23</sup> Overall, these two standard counterexam-

---

<sup>22</sup>Williamson (2009) and Pritchard (2012a, 2016) also defend globalised versions of safety, but there are two important differences between their versions of safety and the proper functionalist version of safety on offer here. Firstly, I motivate globalisation to a set of propositions on the basis of the etiological theory of functions. Neither Pritchard nor Williamson offer a motivation for globalising safety to a set of propositions. Secondly, the proper functionalist criterion of demarcation is more informative than Williamson's and Pritchard's rather vague similarity criterion.

<sup>23</sup>This suggests that relevant modal robustness cannot be equated simply with the possible worlds similar to the actual world where the belief under consideration is true. Since modal robustness is cashed out in terms of proper function, the relevant possible worlds for this safety condition are only those where the belief-forming method and the environment meet a proper functionalist condition specified with reference to the etiological theory of functions. These relevant possible worlds need not be the close similar worlds where the belief is true: the proper functionalist conditions must also be met in the worlds relevant for this

ples to the sufficiency of safety for knowledge fail to land against the new version of safety developed above.

## 5.7 Chapter Summary

I have developed a new version of safety by appeal to proper function and to the etiological theory of functions, and I have explained how this version of the safety condition withstands the most pressing objections to the necessity and sufficiency of safety for knowledge: as a result, it provides also a seemingly successful anti-Gettier condition. I now proceed to compare this new formulation with the more standard versions of safety and with other anti-Gettier conditions more generally. I do so to bring out the distinctiveness of this proper functionalist safety condition and to mark further differences with other prominent modal and non-modal conditions on knowledge.

Let's begin with the standard versions of safety. Unlike Sosa's formulation, the proper functionalist safety condition makes no explicit reference to counterfactuals. Moreover, unlike Pritchard's and Williamson's formulations, it is indexed to both (properly functioning) belief-forming methods and (appropriate) cognitive environments. For all these reasons, it is importantly different from the standard versions of safety.

These differences also distinguish the safety condition on offer from further prominent modal and non-modal conditions on knowledge. Because it doesn't incorporate any ability condition, it diverges from both impure and pure/robust versions of virtue epistemology. And since the concept of normality plays only a background role in comparison with the more central notions of proper function and the etiological theory of functions, this safety condition is different from both modal virtue epistemology and normalcy epistemologies more generally. Recall also that the condition on knowledge developed in this paper is essentially modal: it views knowledge as a matter of (suitably understood) modal robustness across relevant possible worlds. Accordingly, it clearly contrasts with explanationist epistemologies that reject modal conditions on knowledge altogether. Finally, this version of safety is offered as robust anti-Gettier condition: as such, it parts ways also with knowledge-first circular accounts of safety. In general, this version of safety differs from any other extant anti-Gettier condition.

However, just like many previously proposed anti-Gettier conditions, this novel version of safety may still fall prey to Gettier-style counterexamples. There are structural reasons to think this: after all, following Zagzebski (1994), one need only apply a two-step procedure to generate Gettier-style cases against *any* TB + X condition on knowledge, and the proper functionalist safety condition developed in this paper certainly falls under this

---

version of safety. Pavese and Beddor (2018: 69) also divorce relevant possible worlds from close possible worlds. While their ability condition is restricted to possible worlds that are "normal for the task at hand" (Pavese and Beddor 2018: 70) this safety condition is restricted to possible worlds that are normal given the etiological function of the belief-forming method under consideration. Their 'performative' account of normality is thus different from the etiological account of normality endorsed here. For example, facts about history and natural selection in the actual world contribute to determining the modal robustness relevant for this safety condition, but they don't matter for the modal robustness of Pavese and Beddor's skill condition.

category. But, not all is lost: even so, this proper functionalist version of safety remains worthy of serious consideration. Counterexamples are always possible, but whether they are decisive against a purported anti-Gettier condition also always remains to be seen. This proper functionalist version of safety does better than the standard versions of safety: it is more informative, it gains motivation from a number of epistemic and non-epistemic considerations, it avoids the major counterexamples to the necessity and sufficiency of safety for knowledge and it yields the right result in the most discussed Gettier-style vignettes. In light of all these advantages, it would be premature to abandon it just because it may fall prey to further (yet to be raised) counterexamples and objections.<sup>24</sup> As an informative, extensionally adequate and seemingly successful anti-Gettier condition this novel proper functionalist version of safety at least deserves to be taken seriously: absent any such yet-to-be-raised counterexamples and objections, the novel safety condition developed in this paper is still standing.

---

<sup>24</sup>See Weatherson (2003) for a more extensive discussion of this point.

# Chapter 6

## Modalism vs Explanationism: Legal Proof

### 6.1 Overview of the Chapter

In this chapter I introduce and examine the contrast between modalism and explanationism in the context of *applied* legal epistemology. To best do so, I first set the stage by clarifying the key legal concepts routinely discussed in legal epistemology. Along the way, I also explain the subject-matter of legal epistemology, and the points of contact between the study of knowledge and the study of legal evidence and judicial proof. I end by laying out two key case studies at the intersection of epistemology and philosophy of law: the proof paradoxes and the normativity of criminal conviction. While both modalism and explanationism provide a plausible treatment of these case studies, this and the consequent chapters will favour an explanationist – rather than a modalist – applied legal epistemology.

### 6.2 Applied Legal Epistemology

It may seem initially striking to set up *any* kind parallel between epistemology and evidence law. While epistemologists are interested in rather abstract concepts like knowledge, justification and reasons, theorists of evidence law examine the structure and the practicalities of actual trials (preliminary hearing, expert testimony, eye-witness testimony, cross-examination, etc. etc.). Epistemologists develop arguments in favour or against abstract views that are as abstract as the concept they employed, whereas theorists of evidence are concerned with aspects of legal evidence that may make huge practical differences for the verdict of a trial. What is more, epistemologists are interested in setting out conditions for what counts as knowledge or justified belief; by contrast, theorists of evidence of law are concerned with rules that determine what counts as legal evidence. These initial considerations and examples suggest that epistemology – the study of knowledge – and the study of evidence law substantially differ in their subject matters. However,

appearances are misleading, and a closer look reveals many more points of contact between epistemology and evidence law.

Firstly, there's an interesting overlap in the terminology employed by the two disciplines. Here's a few examples of linguistic overlap: "evidence", "reasonable", "doubt", "expert testimony", "eye-witness testimony", "proof", "relevance". A quick glance at the terms used in the two disciplines suggests, as Georgi Gardiner effectively puts it, that legal practice in general – and evidence law in particular – is "up in epistemology to its neck" (Gardiner 2019: 1).

Of course, linguistic evidence is never conclusive. Perhaps these terms mean different things in the epistemology and evidence law contexts, but it's at the very least striking that the concepts and terms of interests of the two disciplines closely resemble each other. However, clear differences between the two domains remain. More specifically, the epistemologist's concept of evidence seems more inclusive than the concept employed by the evidence law theorists. Epistemologists are generous when it comes to enumerate what counts as evidence, and this holds across the board: for example, according to the propositional conceptions of evidence defended by Williamson (2000), a proposition will be part of one's evidence when one knows it regardless of how one came to know said proposition (see Kelly 2016 for discussion). Theorists of evidence law restrict legal evidence to any fact of object that was obtained in compliance with specific evidentiary rules: one's proposition may be part of one's evidence in virtue of being known, and yet, if illegally obtained, will not count as legal evidence. Moreover, legal evidence is evaluated along dimensions that epistemologists do not (and perhaps should not) consider: a belief can be rational even if based on legally inadmissible evidence, and the probative value of a piece of evidence may be much stronger in a trial setting than in a more everyday context. Similar considerations apply to the *sufficiency* of evidence: perhaps what suffices to justify everyday beliefs is not enough when it comes to justify beliefs of guilt and liability. I will return to discuss issues of sufficiency and admissibility in the subsequent sections and chapters; for now it's important to note that despite the initial linguistic overlap, there remains important differences and asymmetries between the legal and epistemological concept of evidence.

A second – and possibly more substantial – point of convergence between epistemology and evidence law theory has to do with the (shared) aim or goal of epistemological inquiry and trials: *truth*. With few exceptions, epistemologists are especially interested in the value of truth: it is a necessary condition on knowledge, it plays key normative roles (see for example truth norms of belief or assertion), it shapes different social institutions and illuminates key social phenomena.<sup>1</sup> Several aspects of society have a clear truth-seeking function, and legal systems make no exception. The point of a trial is to es-

---

<sup>1</sup>The most comprehensive defence of the value of truth in social epistemology is due to Goldman's influential *veritism* (2008) See also (Goldman & O' Connor 2021: Section 2) for discussion. Another influential defence of the value of truth for achieving the aim of criminal justice systems is due to Haack (2014).

establish the truth, and evidence law is one of the many effective ways to achieve this aim.<sup>2</sup> Because of how detrimental it would be for society as a whole if they constantly failed to do so, legal systems show a number of mechanisms to maximize truth-seeking. Traditional epistemology has a key role to play here: it can clarify and illuminate these mechanisms either to describe and better understand legal systems or even to improve them. We may call this second point of convergence “teleological”: perhaps in occasionally different ways, both epistemology and evidence law aim at truth.

Thus, to sum up, given the interesting points of contact between epistemology and evidence law (both linguistic and teleological), applied legal epistemology examines the epistemological issues raised by the concept of legal proof and evidence law. Before actually engaging in applied legal epistemology, in the next section we’ll step back and ask how we can best do so.

### 6.3 Meta-epistemological Desiderata for Legal Epistemology

As we have done already in Chapter 1 in the context of traditional epistemology, before moving on to do applied legal epistemology it may be worth to pause and outline the assumptions that need to be made in order to best engage in it. The reason to do so is two-fold: firstly, it will achieve more clarity in thinking about the intersection between epistemology and evidence law. Secondly, theorising about the epistemic aspects of the law must take into account genuinely legal aspects that are usually outside the scope of traditional epistemology. To keep track of the realities of legal systems it is especially important to get clear on these aspects.

In Chapter 1, we distinguished between different three approaches to the analysis of knowledge: semantic, conceptual and metaphysical. Each approach in traditional epistemology focusses on different aspects of knowledge: the linguistic use of the term, the features of its concept or its more fundamental metaphysical nature. Differently, applied legal epistemology *elucidates* and *clarifies* several distinct dimensions of legal practice. Accordingly, applied legal epistemology is not concerned with providing a semantic, conceptual or metaphysical theory of legal evidence or legal proof. Rather, it attempts to clarify some key legal concepts that have a more or less strong connection to some other notions routinely discussed in epistemology. Here’s two cases in point. For example, the graded notion of legal evidence – its *sufficiency* – mirrors the equally graded notion of justification (beliefs can be *more or less* justified). Accordingly, theories of justification may shed light on the sufficiency of legal evidence.<sup>3</sup> Criminal convictions obey to a de-

---

<sup>2</sup>In *Funk vs United States*, the Supreme Court declared that “The fundamental basis upon which all rules of evidence must rest – if they are to rest upon reason – is their adaptation to the successful development of the truth.” See Ho (2008: Chapter 2) for more statements from actual case law on the role of truth for evidence law theory.

<sup>3</sup>According to the Lockean view of rational belief, a belief is epistemically justified if and only if it’s sufficiently probable. This picture would fit naturally with an elucidation of the sufficiency of legal evidence

manding “beyond reasonable doubt” standard; epistemologists have an awful lot to say about what may or may not make doubts “reasonable”. But of course this *clarificatory* theorising ought to be done always with an eye on the actual legal practice: any theory of reasonable doubt should be in principle available to jurors and juries who are not attuned to or concerned with the epistemology use of the term.

These quick points suggest two important metaepistemological desiderata for applied legal epistemology:

- *(Legally constrained) extensional adequacy.* Applied legal epistemology should make extensionally adequate predictions; that is, epistemological reconstruction of evidence law should be in keeping with what counts as an item of evidence according to contemporary Anglo-American legal systems.
- *Procedural feasibility.* Epistemological theorising within the legal domain should respect procedural constraints: in principle, it could be implemented by contemporary Anglo-American legal systems, and it should not be in tension with the procedural rules of these systems.

With these two desiderata in mind, we can now move on to consider some epistemological aspects of legal proof. This will pave the way for the discussion of the proof paradox.

## 6.4 Legal Proof and Proof-Paradoxes

There are some preliminary and important aspects of doctrine and policy surrounding the process of proof. Whether it is a matter of civil cause for action, serious crimes or affirmative defenses, one party has the initial *burden* to prove each litigated element. This is referred to as “burden of proof”. In the case of a civil trial, it falls on the party that initiates legal action – standardly referred to as “plaintiff”. In the case of a criminal trial, the burden of proof falls on the prosecution. The burden of proof in criminal trials is constrained by the presumption of innocence: because it is worse to convict the innocent rather than letting the guilty go free, the defendant is by default presumed to be innocent until proven otherwise.<sup>4</sup> The presumption of innocence is considered to be a “golden thread” for contemporary legal systems, and it informs the burden of proof in criminal trials.<sup>5</sup>

---

for a rational belief of guilt. See Jackson (2020: Section 3.1) for discussion of the Lockean view.

<sup>4</sup>This asymmetry is often referred to as “Blackstone Ratio”. Blackstone famously (allegedly?) said that “it’s better than that ten guilty defendants go free rather than even just one innocent be convicted”

<sup>5</sup>Statement of the presumption of innocence can be found in the Universal Declaration of Human Rights, Art 11(1) and the Charter of Fundamental Rights of the European Union, Art 48. In the context of the history of Anglo-america Criminal law, It featured prominently in an actual trial, *Woolmington v DPP*, where the it was referred to as the “golden thread”. Viscount Sankey famously said that “Throughout the web of the English Criminal Law one golden thread is always to be seen,that it is the duty of the prosecution to prove the prisoner’s guilt subject ... to the defence of insanity and subject also to any statutory exception. If, at the end of and on the whole of the case, there is a reasonable doubt, created by the evidence given by either the prosecution or the prisoner ... the prosecution has not made out the case and the prisoner is entitled to an acquittal. No matter what the charge or where the trial, the principle that the prosecution must prove

The burden of proof is best broken down into two parts: the burden of production (the duty to present and provide evidence) and the burden of persuasion (the duty to persuade the opposing party of the guilt or liability of the defendant). Crucially, burdens of proof generate distinct *standards of proof*: loosely speaking, standards of proof specify how and when a fact is proven for legal purposes, and legal proof corresponds to the evidence adduced to prove the litigated facts.

Depending on the gravity of the charge and the trial under consideration, we can distinguish three different standards (one increasingly more demanding than the other):

- *Preponderance of evidence*; also called “*More likely than not*”. This standard is applied in the context of relatively low-stakes civil trials (for example: trespassing).
- *Clear and compelling evidence*. Slightly more demanding than the preponderance of evidence standard, this standard is applied in the context of more serious civil trials (for example: adoption cases).
- *Beyond reasonable doubt*. Clearly more demanding than both the preponderance of evidence and clear and compelling evidence standards, this standard is applied in the context of serious criminal trials (for example: murder).

The distinct standards have clear features. They are increasingly more demanding, and as such have a graded nature. They are also essentially connected to (legal) *fact-finding*: they aim to prove and establish what has or has not happened. They apply to specific elements: to prove trespassing in the context of a civil trial each element of trespassing must be proven (actual trespassing *and* intention to trespass). Similar considerations can be made regarding other civil and criminal trials.

However, the details are less clear, and trying to get clear on them raises a host of philosophical – more specifically, epistemological – questions. What is it to *prove* a fact for legally acceptable standards? Is it to *justifiably believe* it or to *know* it? How does this difference carry over to more or less demanding standards? How is the risk of false convictions to be managed? How is it to be understood?

Epistemological questions pertaining to the nature of standards of proof and legal proof are made especially vivid by a number of seemingly intractable paradoxes discussed by theorists of evidence law. One (once very popular) way to make progress on the questions raised above consists in adopting a probabilistic interpretation of the standards. According to the doctrine of legal probabilism, standards of proof can be reduced to suitably specified probabilistic thresholds.<sup>6</sup> Accordingly, legal proof is a matter of sufficiently high probabilistic support depending on the standard under consideration.

---

the guilt of the prisoner is part of the common law of England and no attempt to whittle it down can be entertained”. See Picinali (2021) for a recent discussion of the value and significance of the presumption of innocence.

<sup>6</sup>See Di Bello and Urbaniak (2021) for an updated overview. See also Ho (2008: Chapter 4; 2021: Section 3.2) for discussion. I return to legal probabilism in more detail in the next chapter.

Legal probabilism has virtues. It provides a principled and informative elucidation of the standards and it captures their *graded* nature. However, legal probabilism generates well-known paradoxes in evidence law. These paradoxes involve *merely statistical* evidence and are made vivid by the following stylised court cases:

**Blue Bus.**<sup>7</sup> While driving back home, a vehicle suddenly hits Jane’s car. In the accident, she breaks her leg. She realises that the vehicle was a bus, but she cannot discern the colour. Jane runs a quick search, and discovers that the Blue Bus Company operates 80% of buses in town. The remaining 20% are operated by the Red Bus Company. Jane sues the Blue Bus Company. Using only the evidence described here, Jane expects to win the case. After all, given the evidence, it’s 0.8 probable that the Blue Bus Company is liable.

**Prisoners.**<sup>8</sup> 100 prisoners are exercising in the prison yard. Suddenly 99 of them attack the guard, putting into action a plan that the 100th prisoner knew nothing about. The 100th prisoner played no role in the assault and could have done nothing to stop it. There is no further information that can be used to settle the question of any particular prisoner’s involvement. Despite the paucity of evidence, the guards pick prisoner 69 and convict him for the attack. After all, given the evidence, it’s 0.99 probable that prisoner 69 is guilty.

These hypothetical cases suggest that no matter how high the probability of liability or guilt is, it still remains problematic and unfair to impose liability or guilt on the Blue Bus company or one prisoner. These cases are commonly referred to as “proof paradoxical” scenarios (2020b), and they put pressure on a probabilistic foundation of legal proof.

These are not merely hypothetical cases: examples like these abound also in actual case law. For example, **Blue Bus** closely mirrors the civil lawsuit filed by Mrs. Smith against Rapid Transit, a small bus company in Massachusetts. Here’s a more detailed reconstruction of the case:

*Smith v. Rapid Transit, Inc.* At around 1:00 A.M. on Feb. 6, 1941, while driving an automobile on Main Street in the City of Winthrop, plaintiff Smith she observed a bus coming toward her, which she described as a “great big, long, wide affair”. The bus, which was owned and operated by defendant Rapid Transit, Inc. (“Rapid”), was proceeding at about 40 miles an hour when, according to Smith, it forced her to turn to the right, causing her automobile to collide with a parked car. The department of public utilities had issued a certificate of public convenience or necessity to Rapid for three routes in Winthrop, one of which included Main Street. However, another bus line was in operation in Winthrop at that time of the accident, but not on Main Street.

---

<sup>7</sup>Redmayne (2008), Enoch et al. (2012), Smith (2018), Gardiner (2019) discuss this case.

<sup>8</sup>Originally described in Nesson (1979b), here I’m adopting the presentation from Redmayne (2008).

Smith filed a lawsuit in Massachusetts commonwealth court seeking damages for injuries she sustained in the accident. Smith claimed that Rapid's employee was negligent in operating the bus, which caused the accident.<sup>9</sup>

Perhaps surprisingly for the advocates of legal probabilism, the court refused to impose liability on Rapid. The words of the negative verdict returned by the court are especially telling. The liability of Rapid was “a matter of conjecture”, and the “mathematical chances” were *not enough* to prove it.

Statistical evidence is not foreign to actual criminal law either. For example, the use of DNA evidence in criminal trials is becoming increasingly popular, and it led to valuable exonerations for wrongly convicted citizens.<sup>10</sup> Conversely, the use of merely statistical evidence has also led to problematic wrongful convictions. Lucia de Berk, a Dutch paediatric nurse, was wrongfully condemned to life imprisonment on the basis of (incorrect) statistical calculations. Similarly, a couple in California was wrongfully accused of murder just because they were thought to match a witness description on the basis of incorrect statistical estimations.<sup>11</sup> Overall, these examples of actual case law suggest the virtues (and the limits) of legal probabilism.

Applied legal epistemology can offer useful tools to solve the proof paradoxes and offer different understandings of legal proof. Indeed, the two views discussed in the previous chapters – modalism and explanationism – can be applied to both analyse and elucidate *both* knowledge and legal proof. Counting also the probabilistic interpretation, we may enrich the options by adding both modalism and explanationism:

- **Probabilistic standards of proof.** Legal proof is a matter of suitably specified *probabilistic support*: legal verdicts must be *likely* to be true.
- **Modalist standards of proof.** Legal proof is a matter of suitably specified *modal robustness*: given the evidence adduced, legal verdicts track the truth across relevantly close possible worlds (given the evidence upon which they are based, they continue to be true).
- **Explanationist standards of proof.** Legal proof is a matter of suitably specified explanatory connection. Legal proof must meet explanatory thresholds: only sufficiently good explanations of the evidence adduced can meet the standard under consideration.

The modalist and explanationist accounts of the standards closely mirror the modalist and explanationist analyses of knowledge laid out and discussed in the previous chapters. The relationship between knowledge and legal proof is however delicate: while both knowledge and legal proof display a relevant connection to truth, it may strike as odd to simply

---

<sup>9</sup>Available at <https://www.lexisnexis.com/community/casebrief/p/casebrief-smith-v-rapid-transit-inc>.

<sup>10</sup>See Roth (2010) for an interesting discussion of the use of DNA evidence in actual case law. I return to this in more detail in the next chapter.

<sup>11</sup>The legal case is *People v. Collins*.

identity knowledge and legal proof. I return to the delicate relationship between knowledge and legal proof in the next two chapters; in particular, I will look at the relationship between knowledge, legal proof and epistemic luck. For present purposes, it's important to note that modalism and explanationism can provide the conceptual resources to better understand the nature of legal proof.

Here's an initial worry about the application of modalism and explanationism to legal epistemology. One may worry that the *graded* notion of legal proof can't be fully captured by *all-or-nothing* modalist and explanationist conditions on knowledge, and that a probabilistic view is better suited. However, both modalism and explanationism can capture the graded notion of legal proof. Start with modalism: as we have seen in Chapter 2, a safety condition on knowledge can be weak or strong depending on the number of close worlds where the belief must continue to be true. Advocates of modalist accounts of legal proof can exploit this degree of demandingness to capture a weaker civil standard and a stronger criminal standards. Similar considerations apply to explanationism: better explanations may be required to meet more demanding standards.

And here's also an initial advantage in applying modalism and explanationism to legal epistemology. In different ways, modalism and explanationism provide a diagnosis of the proof paradox. Merely statistical evidence fails to meet a modalist standard of proof: much like a lottery-style proposition, because the verdict is based on merely statistical evidence it could easily be false (in close worlds, it could too easily be mistaken). Merely statistical evidence also fails to meet an explanationist standard of proof: because of its lack of connection with the facts at issue, the evidence doesn't really explain why the defendant is guilty or liable (more on this in the next chapters).

Let's take stock. Interestingly, modalism and explanationism can be applied to make progress on various epistemological questions underlying the nature of legal proof. Before considering and assessing specific versions of what we may call "legal modalism" and *legal explanationism*, I need to clarify how to best approach the proof paradoxes and lay out two case studies for legal modalism and explanationism.

## 6.5 Explanatory, Descriptive and Normative Projects

I have already hinted at how modalism and explanationist interpretations of standards of proof can diagnose what goes wrong with the proof-paradoxical scenarios. However, because we are trying to provide an answer which displays *procedural feasibility* and (legally constrained) *extensional adequacy*, it is worth being more precise about these diagnoses of the proof paradox. Here's a number of *distinct* questions we can ask by approaching the proof-paradox and while theorising about the epistemology of legal standards of proof:

- What do our (*lay*) intuitions say about the paradoxes? (Question 1)
- How do our intuitions diverge from the intuition of *trained jurors, judges and legal actors*? (Question 2)

- Can legal epistemology *capture* how legal practice actually works? (Question 3)
- Can legal epistemology inform how legal practice *should* ideally work? (Question 4)

These are all different questions, and there's no reason to assume that we can give one straightforward consistent answer to each.<sup>12</sup> Following Fratantonio's insightful taxonomy (Fratantonio 2021), we can distinguish at least three independent approaches to the proof-paradox:

- The **explanatory** project.
- The **descriptive** project.
- The **normative** project.

The explanatory project diagnoses the intuitions of laypeople and legally trained legal actors in reaction to the proof-paradoxical scenarios (Question 1 and 2 above). The descriptive project aims at capturing and explicating how legal proof works (Question 3 above). The normative project, more ambitiously, is supposed to theorise about how legal proof should work (Question 4 above).

Crucially, modalism and explanationism can be applied and assessed in the context of each of these projects. A safety approach to legal proof may diagnose the conflicting intuitions of laypeople and legal actors towards statistical evidence (explanatory project). It may capture how the law operates (descriptive project), or inform how the law should operate (normative project). The same goes for an explanationist approach to the standards.

Crucially, emphasising the independence between the projects doesn't mean that they can't be pursued together. Indeed, one may start out with the explanatory project and then move on to the descriptive project: there is an independently plausible connection between how the law operates and legal actors' intuitions. One may as well take the intuitions against the use of statistical evidence as a symptom of some epistemic impropriety, and then identify a normative aspect of the standards of the proof that statistical evidence fails to meet. Carefully distinguishing the three projects does not preclude the possibility of pursuing them together.

With these points in play, we can now move on to consider more carefully two important case studies in applied legal epistemology: the epistemic quality of individualised evidence and the epistemic normativity of criminal convictions.

### 6.5.1 Case Study I: Individualised versus Statistical Evidence

The notion of "individualised evidence" represents a hugely debated legal category. Precise definitions of the notion are however scarce, and it's best introduced by means of

---

<sup>12</sup>These questions are implicitly raised and attended in virtually any recent legal epistemology paper. These questions are explicitly distinguished in Spotswood (2021).

examples from more or less realistic legal cases. It's useful to think of individualised evidence as the counterpart of merely statistical evidence: crucially, individualised evidence displays a direct connection with the facts at issue. This is not the case with merely statistical evidence, which simply makes guilt or liability highly probable. Drawing on the proof-paradoxical scenarios outlined above, here's some examples of commonly discussed instances of individualised evidence:

**Eye-witness testimony.** A guard reports that they saw the prisoner attacking the guard, or an eye-witness reports that they saw a blue bus hitting the car.

**Video-evidence.** Recorded images of a prisoner attacking the guard, or of the blue bus hitting the car.

As it is clear from these examples, the evidence under consideration displays some sort of *direct* connection with the fact at issue.<sup>13</sup> However, legal theorists and evidence law theorists have often struggled to identify in details what makes evidence individualised to the facts at issue.<sup>14</sup> A potentially fruitful approach to this problem employs the two frameworks discussed in this dissertation. For example, individualised evidence can be characterised by appeal to a modalist or explanationist necessary condition which identifies the epistemic quality of this much debated type of legal evidence.<sup>15</sup> Accordingly, we get:

- *Modalist individualised evidence.* A piece of evidence *e* qualifies as individualised evidence *only if* it is *safe* (or *sensitive*).
- *Explanationist individualised evidence.* A piece of evidence *e* qualifies as individualised evidence *only if* it displays appropriate explanatory connections with the facts at issue.

In the subsequent chapters, both approaches will be discussed and assessed.

---

<sup>13</sup>A potentially more controversial instance of individualised evidence is DNA evidence, and in this chapter I will remain non-committal about whether DNA evidence counts as individualised. However, I want to make the following two observations to explain to the reader why this issue is especially complicated. First, like other instances of statistical evidence, DNA evidence is presented in court by an expert in statistical form: basically, an expert testifies about the (usually) stupendously low probability of a random match between the DNA profile found on the scene and the suspect's. Secondly, like other instances of individualised evidence, correct DNA matches are individualised in the sense that they refer to specific individuals (they are, in some sense, about the suspect who may be substantially involved with the facts at issue). One final word of caution: unlike other instances of individualised evidence (e.g., eye-witness testimony), DNA matches should not be equated with guilt even when correct: a correct eye-witness identification somehow naturally implies guilt (e.g., the witness correctly reports that they saw the defendant committing the crime), but a correct DNA match suggests *at best* the presence of the defendant on the scene.

<sup>14</sup>See Spootswood (2021: ) for an exhaustive list of references.

<sup>15</sup>See Redmayne (2008: Section ) Dahlman & Pundit (2021) for arguments to the effect that the quality of individualised evidence is *moral* as opposed to *epistemic*.

## 6.5.2 Case Study II: The Epistemic Normativity of Criminal Conviction

Another relevant case study in applied legal epistemology concerns the epistemic normativity of criminal convictions. When epistemologists talk about norms they have in mind some sort of *standard of appropriateness* for the epistemic dimension of an activity with an aim. Take for instance the extensive literature on the norm of assertion: each proposed norm purports to capture the genuinely epistemic propriety of the speech act of assertion.<sup>16</sup> Similar considerations apply in the contexts of criminal verdicts: which norm, if any, can capture the epistemic normativity of criminal convictions? This purported norm is suitably restricted to the epistemic propriety of criminal convictions, and it's meant to sidestep issues concerning procedural, legal, moral or practical aspects of criminal convictions. Jennifer Lackey nicely makes this point as follows:

The kind of propriety that is at issue here is epistemic in nature. The norm captures what status jurors need to have with respect to the question of the defendant's guilt in order to be free from distinctively epistemic criticism when registering a guilty verdict. There may be other sorts of criticism that are appropriate if one violates this norm, such as moral, legal, and so on. But if this norm is correct, then when, and only when, jurors have knowledge is it epistemically appropriate for them to convict. (Lackey 2021: 1,2)

Options to capture the epistemic propriety of criminal convictions abound. Once we conceive of criminal convictions as constituted by a suitable combination of speech acts (the jury *declaring* the defendant guilty) and actions (the punishment *inflicted* on the guilty defendant), the same norms governing speech acts and actions can be applied to capture the normativity of criminal convictions. For example:

**Knowledge Norm of Criminal Conviction.** Convict a defendant if and only if you know he is guilty.

**Truth Norm of Criminal Conviction.** Convict a defendant if and only if he is guilty

**Justification Norm of Criminal Conviction.** Convict a defendant if and only if you justifiably believe he is guilty.

As it should be clear, similar norms have been put forth to capture the epistemic propriety of assertion.<sup>17</sup> We can tell a similar story in the context of criminal convictions. Suppose that a judge convicts the defendant on the basis of sheer prejudice, or by randomly flipping a coin.<sup>18</sup> Clearly, these convictions are not appropriate from an epistemic point of view.

<sup>16</sup>For a nice overview, see the supplement in Pagin and Marsili useful survey piece (Pagin and Marsili: 2021).

<sup>17</sup>Defenders of the knowledge norm of assertion include Unger (1968), 2000, DeRose (2002, Hawthorne (2005), Engel (2008), Schaffer (2008), Sosa (2015 Simion (2016a).

<sup>18</sup>Judith Thomson (1986) famously considers this case; I return to this in detail in the next chapter.

Or further suppose that the jury is Gettierised, and their true belief of guilty and liability is also true as a matter of luck or coincidence (this verdict is based on a justified true belief that yet fails to constitute knowledge).<sup>19</sup> Or maybe the jury carefully considers the evidence, reaches a verdict, but for an abnormal twist of event the verdict ends up being false (this verdict is based on a justified false belief). Perhaps more controversially, these verdicts are also epistemically improper, and different norms of convictions explain why. The justification norm diagnoses the problem with a conviction based on prejudice and a coin toss (the beliefs under consideration are not justified). The knowledge norm diagnoses Gettierised convictions, and the truth norm does the same with *any* wrongful convictions.

At precisely this point, modalism and explanationism enter the picture. For example, suppose that, following a promising modalist approach, knowledge equates with a type of safe belief. Then the knowledge norm of criminal conviction will require a safe belief as necessary and sufficient for appropriate convictions. Similarly, suppose that justified belief requires some kind of explanatory connection with the proposition believed: then the justification norm will be unpacked along explanationist lines. Back to the examples above: convictions based on prejudice do not track the truth in close possible worlds and fail to display the relevant explanatory connection to the fact at issue. Gettierised convictions fail to meet some type of modal or explanationist condition on knowledge.

Once again, both modalism and explanationism provide the needed conceptual resources to articulate purported norms of criminal convictions. If we grant the criminal convictions have a genuine epistemic dimension, then we can move on to develop a norm that captures such dimension along modalist or explanationist lines.

## 6.6 Swapping: Against Modalism and For Explanationism

With these points in play, here's a road map for the upcoming chapters. While I have defended modalism and objected to explanationism in the context of traditional epistemology, I will now do the reverse: when it comes to applied legal epistemology, I maintain that explanationism (rather than modalism) holds the key to yield a successful account of individualised evidence, the epistemic normativity of criminal convictions and judicial proof more generally. This is because knowledge and judicial proof are *structurally* different, and require two distinct approaches (I will make sure point out these important structural differences along the way). Accordingly, here's a more detailed overview of the remaining chapters.

In Chapter 7, I move on to make a rather negative point: I develop novel objections against currently popular modal accounts of individualised evidence. Focussing on anti-luck condition, I point out important differences between knowledge and individualised

---

<sup>19</sup>For example, Pardo (2010) equates legal proof with a non-Gettierised true belief.

evidence, and I show how neither safety nor sensitivity can capture the legal notion of individualised evidence.

In Chapter 8, I make a more constructive proposal: I diagnose the failure of safety and sensitivity characterisations of legal evidence and provide a different explanationist account. I offer two arguments in its support; along the way, I also sketch out important consequences of explanatory connections for the epistemic normativity of criminal convictions. The final upshot is a well-informed account of legal proof which does away with modal conditions and better captures the nature of juridical proof.

## **6.7 Chapter Summary**

This chapter has introduced some key issues of applied legal epistemology. I have noted some analogies and disanalogies with traditional epistemology, and looked at two case studies (individualised evidence and the epistemic normativity of criminal convictions) through the lens of modalism and explanationism.

# Chapter 7

## Against Modal Accounts of Legal Proof

### 7.1 Overview of the Chapter

The notion of individualised evidence holds the key to solve the puzzle of statistical evidence, but there's still no consensus on how exactly to define it. To make progress on the problem, epistemologists have proposed various accounts of individualised evidence in terms of causal or modal anti-luck conditions on knowledge like appropriate causation (Thomson 1986), sensitivity (Enoch et al. 2012) and safety (Pritchard 2018b). In this chapter, I show that each of these fails as satisfactory anti-luck condition, and that such failure lends abductive support to the following conclusion: once the familiar anti-luck intuition on knowledge is extended to individualised evidence, no single causal or modal anti-luck condition on knowledge can succeed as the right anti-luck condition on individualised evidence. This conclusion casts serious doubts on the fruitfulness of the move from anti-luck conditions on knowledge to anti-luck conditions on individualised evidence. I expand on these doubts and point out further aspects where epistemology and the law come apart: epistemic anti-luck conditions on knowledge do not adequately characterise the legal notion of individualised evidence.

### 7.2 Individualised Evidence and Luck

A witness, a tape, a video, a trace: this is evidence *individualised* to a fact. It's not very reliable: witnesses lie or err, tapes can be faked, images are ambiguous, and traces mislead. Statistics, percentages, estimations of probability: this is *bare* statistical evidence. It seems more reliable: it can make crucial incriminating facts highly probable, and seldom leads astray. Yet, the law appears to prefer the former over the latter. But why? What's so special about individualised evidence? These are the main questions raised by the puzzle of statistical evidence, and while the notion of individualised evidence may hold the key to solving it, there's still no agreement on how exactly to define it. In this passage, Judith Thomson gives us a crucial hint:

Suppose that a jury is puzzled by the evidence which has been presented to it, and cannot arrive at a consensus as to its weight. “I know”, says one juror, “let’s decide by flipping a coin: heads we impose liability, tails we don’t.” They agree, and they flip a coin, which comes up heads. So they return and say, “the defendant is guilty.” *Their so doing is not made acceptable by the fact (supposing it a fact) that the defendant actually is guilty.* (...) It matters to us, not just that a defendant does not suffer a penalty unjustly, but also that the penalty is not imposed on him unjustly. (Thomson 1986: 213, emphasis added)

Her remarks are significant, and suggestive. Justice demands that punishment and liability are imposed on the basis of the right evidence, and the right evidence is incompatible with (a certain type of) luck. It’s thus an important desideratum that individualised evidence satisfies an anti-luck condition. The point is familiar to epistemologists: an identical desideratum applies to knowledge. Indeed, the main challenge underlying the Gettier problem consists in identifying the right anti-luck condition on knowledge, as Duncan Pritchard (2015: 94) points out here:

Gettier’s famous chapter didn’t just demonstrate that knowledge wasn’t (non-factively) justified true belief. He also demonstrated that one could have a justified true belief which was nonetheless subject to knowledge-undermining epistemic luck. This raises a challenge regarding what condition or conditions must be imposed on knowledge in order to exclude such luck. Call this formulation of the Gettier problem the anti-luck problem.

The anti-luck problem seems to affect both knowledge and individualised evidence, and indicates a deeper structural connection between the two. This connection did not go entirely unnoticed: epistemologists have offered accounts of individualised evidence in terms of epistemic qualities which are also advanced as anti-luck conditions on knowledge. However, when imposed on knowledge, these conditions are unsuccessful: they face perfectly general and possibly unavoidable difficulties. In this chapter, I take up the task to show that the same difficulties carry over to individualised evidence: my conclusion will be that, given the difficulties incurred with knowledge, no single causal or modal anti-luck condition on knowledge can succeed as the right anti-luck condition on individualised evidence. And this is a potentially significant result, for it casts into serious doubt the fruitfulness of the move from an anti-luck condition on knowledge to an anti-luck condition on individualised evidence.

My plan is as follows. In §1, I set the stage and lay out the puzzle in detail. In §2, I clarify further the connection between knowledge, individualised evidence and luck. Next, in §3 and §4, I offer Gettier-style variations of stylized court cases and show that prominent causal and modal anti-luck conditions on knowledge fail also when imposed on individualised evidence. Finally, in §5, I advance a diagnosis of the failure of these prominent anti-luck conditions. I expand on the relevance of such failure and I point out key aspects in which epistemology and the law come apart; this has important repercussions for legal epistemology in general.

### 7.3 The Puzzle of Statistical Evidence

In a fair trial, defendants are not required to prove their innocence: the onus of proof always falls on the party who first takes legal action. Such asymmetry stems from what is considered the golden thread of many contemporary criminal justice systems, the presumption of innocence. The *burden of proof* is so established: it is a necessary condition for conviction that the State proves the guilt of the defendant, who famously remains innocent “until proven guilty”.<sup>1</sup>

The burden of proof generates a *standard of proof*. While the current standard for civil liability is the “preponderance of evidence”, the standard for criminal conviction is the more demanding proof “beyond reasonable doubt”. These formulations are renowned, but their exact meaning is far from clear. Actually, it appears to be surprisingly *elusive*: explaining the precise nature of the standards is no easy task even for trained legal officials. As chief judge Jon Newman (1993: 984) effectively put it, “I find it rather unsettling that we use a formulation that we believe will become less clear the more we explain it”.

Among legal theorists, it is sometimes assumed that probability theory can provide a way around this conundrum. A once very popular view in legal scholarship favours a *quantitative* interpretation: roughly, the civil and criminal standards neatly reduce to quantifiable *probabilistic thresholds*.<sup>2</sup> Following Blome-Tillman (2017), they can be defined in terms of evidential probabilities:

**Preponderance of Evidence:** Proposition  $p$  meets the standard iff  $P(p|e) > 0.5$

**Beyond Reasonable Doubt:** Proposition  $p$  meets the standard iff  $P(p|e) \geq 0.9$

So understood, the preponderance of evidence and beyond reasonable doubt standards are satisfied so long as, given the admissible evidence adduced to the case, the probability of liability and guilt is at least 0.6 or 0.9 respectively. The key motivation for this probabilistic interpretation is to ensure the accuracy of legal verdicts and secure the Blackstone ratio, according to which the percentage of false convictions should be significantly lower than false acquittals.

So far, so good. However, this probabilistic interpretation leads to a puzzle long discussed in evidence law – the puzzle of statistical evidence.<sup>3</sup> Let’s consider the following two cases, ranging over the civil and criminal standards:

---

<sup>1</sup>Scots law makes an exception, and permits a third verdict (“not proven”) in addition to “guilty” or “not guilty”. If this possibility obtains, then the defendant does not remain innocent until proven guilty.

<sup>2</sup>Using Haack’s established terminology, this view goes under the name of legal probabilism and it has both prominent defenders (e.g., Calabresi 1961, Becker 1961, Posner 1973, Hedden and Colyvan 2019) and prominent critics (e.g., Tribe 1971, Cohen 1977, Allen and Leiter ?, Allen and Stein 2013, Haack 2014). See Di Bello and Urbaniak (2021) for a thorough and updated discussion of legal probabilism. The “once” proviso is especially important: these days, theorists of evidence law have moved away from legal probabilism and endorsed an explanationist interpretation of legal proof in terms of relative plausibility. See Allen (2008), Pardo and Allen (2008, 2019) for an exhaustive presentation of an explanationist account of the main standards of proof.

<sup>3</sup>The puzzle is also known as the proof paradox. See Gardiner (2019; forthcoming) Ross (2020b), Pardo (2019: §3), Ho (2008: 135-142) and Redmayne (2008: §1) for an overview.

**Accident.**<sup>4</sup> While driving back home, a vehicle suddenly hits Jane’s car. In the accident, she breaks her leg. She realises that the vehicle was a bus, but she cannot discern the colour. Jane runs a quick search, and discovers that the Blue Bus Company operates 80% of buses in town. The remaining 20% are operated by the Red Bus Company. Jane sues the Blue Bus Company. Using only the evidence described here, Jane expects to win the case. After all, given the evidence, it’s 0.8 probable that the Blue Bus Company is liable.

**Prisoners.**<sup>5</sup> 100 prisoners are exercising in the prison yard. Suddenly 99 of them attack the guard, putting into action a plan that the 100th prisoner knew nothing about. The 100th prisoner played no role in the assault and could have done nothing to stop it. There is no further information that can be used to settle the question of any particular prisoner’s involvement. Despite the paucity of evidence, the guards pick prisoner 54 and convict him for the attack. After all, given the evidence, it’s 0.99 probable that prisoner 54 is guilty.

A puzzle looms ahead. While the evidence presented in statistical form meets the probabilistic standards, it also seems insufficient to undergird a positive legal verdict. In the vignettes above, imposing liability on the Blue Bus Company and convicting prisoner 54 seems unjust.<sup>6</sup> Conversely, a different type of evidence - individualised evidence - changes our reaction to the cases. The notion of individualised evidence takes centre stage in this chapter, but it has proven surprisingly resistant to a precise definition (I return to this in §3). Hence, at this preliminary stage I must appeal to canonical examples to make it clearer.<sup>7</sup> Suppose that in **Accident**, Jane’s usually reliable and trustworthy friend Jack *tells* her that he saw a blue bus hitting her car. Or imagine someone *seeing* prisoner 54 assaulting the guard. Further suppose that, after running several tests, it turns out that the eye-witnesses provide accurate identifications only 70% of the time in similar lighting conditions. On such basis, verdicts of liability and guilt are more acceptable, but neither piece of individualised evidence is as probative as the market share of the two companies, or the stupendously high probability that prisoner 54 assaulted the guard. As respectable psychological studies show, eye-witnesses are only imperfectly reliable.<sup>8</sup>

---

<sup>4</sup>Redmayne (2008), Enoch et al. (2012), Smith (2018), Gardiner (2019) discuss this case.

<sup>5</sup>Originally described in Nesson (1979b), here I’m adopting the presentation from Redmayne (2008).

<sup>6</sup>The use of toy-cases like these may muddy the waters. As Steele and Colyvan (2022) emphasise, the presentation in **Prisoners** is problematic because the intuitions of unfairness that it prompts track more equity issues rather than the use of statistical evidence: since each prisoner has the same probability of being guilty, it’s unfair to randomly single out and punish only one prisoner. To better appreciate the point, imagine a line of cars that are all illegally parked. A parking inspector arrives, sees and knows that they are illegally parked but fines only one of the owners. This is unfair because of inequity: one car owner is unreasonably singled out (thanks to a referee for *Philosophical Studies* for raising the case). Perhaps the **Prisoners** case can be changed so that each of the 100 prisoners is punished, but then it’s no longer clear that securing 99 known to be correct convictions is unfair (many would take this to be a very good outcome). These observations bring out the complexity underlying these toy-cases and anticipate the problems with using them to advance the debate on statistical evidence (more on this in §5).

<sup>7</sup>See Enoch and Fisher (2015: §1) and Pardo (2019: 262-266) for discussion of extant attempts to distinguish statistical from individualised evidence.

<sup>8</sup>Kahneman and Tversky (1982), Coady (1992: Chapter 15), Loftus (1996; 2019: §3).

The puzzle gets closer. On the basis of canonical pieces of individualised evidence, the probability of liability and guilt is lower than it would be if only statistical evidence was adduced. Individualised evidence would even fail to meet the probabilistic threshold required for criminal conviction: given the reliability of eye-witnesses, it is only (roughly) 0.7 probable that prisoner 54 attacked the guard. Yet such evidence seems better suited than mere estimations of probability for securing a positive verdict. This squares badly with the probabilistic interpretation of the standards: why do we intuitively eschew highly reliable statistical evidence but accept imperfectly reliable individualised evidence?

This question is neither just a matter of subjective intuitions nor an instance of abstract philosophical speculation: legal practice also shows a rather conflicted attitude towards purely statistical evidence.<sup>9</sup> **Accident** is very similar (almost identical) to an actual civil lawsuit, *Smith v. Rapid Transit Inc.* Unsurprisingly, the bus company was *not* found liable. Moreover, studies involving mock juries suggest that trained judges are equally reluctant to impose liability using statistical evidence alone, and also have a preference for individualised evidence (e.g., Wells 1992).<sup>10</sup> Such antipathy is best appreciated in terms of *sufficiency* rather than *admissibility*: statistical evidence is not inadmissible, but legal officials seem hesitant to convict solely on the basis of it.<sup>11</sup>

The puzzle is here. On the one hand, according to a probabilistic interpretation, statistical evidence should be sufficient to meet the standards, but both intuitions and legal practice deny this. On the other, according to a probabilistic interpretation, individualised evidence should often be insufficient to meet the standards, but both intuitions and legal practice deny this too. We face a puzzle: the puzzle of statistical evidence.

---

<sup>9</sup>This is an empirical observation about the law, so I must proceed carefully. Legal practice is usually averse to *certain* pieces of statistical evidence, but not to others, such as DNA evidence. DNA evidence is presented in statistical form, and yet it has helped to secure many post-conviction DNA exonerations. See Roth (2010) and Ross (2021b: §3) for an illuminating discussion of the use of DNA evidence in trials; consult this page to learn more about post-conviction DNA exonerations <https://www.innocenceproject.org/about/>.

<sup>10</sup>The generally widespread convergence of intuitions against statistical evidence and in favour of individualised evidence may be telling but it's far from decisive. An experimental study conducted by Ebert et al. (2018) suggests that laypeople attribute a positive epistemic status to lottery-style propositions supported by sufficiently high statistical evidence. Moreover, the preference for individualised evidence can be explained by appeal to cognitive bias: Pennington and Hastie's (1991) story model of trial predicts that jurors' decision making is guided by attempts to impose a narrative structure to the evidence presented in court. Crucially, the informational richness of individualised evidence (e.g., eye-witness testimony) displays the exact narrative structure that could explain why it's usually preferred to individualised evidence (see Ross 2020a for helpful discussion). I offer these remarks to weaken the weight that is sometimes attributed to people's intuitions against statistical evidence.

<sup>11</sup>Another empirical observation about the law that warrants caution. The point on sufficiency doesn't always apply: in some key civil cases like toxic torts, statistical evidence is deemed sufficient. This is made especially vivid by the doctrine of market share liability (see *Sindell v. Abbott Laboratories* for a *locus classicus*). Suppose that a defective product causes harm to the plaintiff and it's hard to establish which of the many manufacturers who sell it did in fact sell the defective product to the plaintiff. In this case, each manufacturer selling the product has to reimburse the plaintiff according to the market share which, crucially, is determined by statistics. See Ross (2021b: 9; 2021a) and Enoch and Fisher (2015: 562) for an illuminating discussion of the genuine doctrinal complications with the use of statistical evidence in actual case law.

## 7.4 A Common Anti-luck Condition?

The puzzle has been long discussed in evidence law, but it has recently sparked a lively debate also in epistemology.<sup>12</sup> On the negative side, legal epistemologists have rejected a probabilistic interpretation of the standards.<sup>13</sup> More positively, they have proposed to identify individualised evidence in light of specific epistemic qualities usually imposed as necessary conditions. At this point in the picture, enter knowledge: crucially, the epistemic qualities imposed as necessary conditions on individualised evidence are posited *also* as anti-luck conditions on knowledge. In this chapter, I am not providing a novel solution to the puzzle. Rather, I shall argue that there can't be one in terms of a single anti-luck condition on knowledge. But to do so, I first have to clarify the link between knowledge, luck and individualised evidence.

At closer inspection, such link is not surprising: knowledge and individualised evidence share several common features. Knowledge is factive, so let's start with truth. Perhaps controversially, we can assume that individualised evidence also retains an essential connection to truth: it's propositional and consists of true propositions. This is because one key subproject in legal epistemology (though not the only one) is a *normative* project concerned with the type of evidence that an *ideal* legal system would employ.<sup>14</sup> Once we appreciate such normative dimension and concede that accounts of individualised evidence are cast as normative theses, it should be easy to appreciate also that an ideal system would not (and should not) base convictions on misleading evidence.<sup>15</sup>

Like knowledge, individualised evidence requires more than truth. The point is familiar: it is an important datum in epistemology that knowledge and luck are essentially incompatible.<sup>16</sup> The same holds in the law: legal verdicts which are correct only as a result of luck violate fundamental principles of justice. To appreciate why, recall the thought experiment offered by Judith Thomson (1986: 213). Imagine an incredibly lucky jury that

---

<sup>12</sup>Thomson (1986), Enoch et al. (2012), Buchak (2014) Smith (2016: Chapter 2, 2018), Blome-Tillman (2015, 2017), Pritchard (2018b), Littlejohn (Littlejohn), Moss (forthcoming, 2021), Backes (2020), Gardiner (2019, 2020), Ross (2020a, 2020b, 2021a, 2021b), Di Bello (2019), Fratantonio (2021). It's worth noting that the puzzle has also received a decision-theoretic discussion in Regan et al. (2001).

<sup>13</sup>Hedden and Colyvan (2019) offer a contrary view.

<sup>14</sup>Some legal epistemologists explicitly engage in a normative project (e.g., Hedden and Colyvan 2019: 449, 458). Others are more interested in the descriptive enterprise of interpreting existing legal practice, and prefer to remain neutral on the potentially revisionary consequences of their views (e.g., Smith 2018: 1200).

<sup>15</sup>I must add an important point of clarification. This paragraph won't be palatable to those who reject a factive standard, and see accommodating the possibility of (procedurally correct) wrongful convictions as a virtue (Smith 2018: 1202; Gardiner forthcoming: 7). Perhaps more sophisticated versions of factive standards can be reconciled with the possibility of wrongful convictions (e.g., Blome-Tillman 2017), but a full elaboration of this claim exceeds the scope of the chapter. To the unconvinced reader, I want to offer the following remarks. First: the necessary conditions discussed in this chapter *inevitably* imply truth. Causal relations do not obtain without causal relata, sensitivity implies factivity (false beliefs are never sensitive), and so does the counterfactual formulation of safety originally offered by Sosa (1999: 146), according to which if S were to believe *p*, then *p* would be true ( $B(p) \rightarrow p$ ). Second: as the opening quote from Thomson shows, we have good independent reasons to think that truth, even if necessary, is ultimately insufficient for individualised evidence (hence the need for an anti-luck condition). In light of these points and in order for my discussion to get off the ground, I ask the reader to charitably concede that individualised evidence requires truth.

<sup>16</sup>See Unger (1968) for early work on luck, and Carter and Berrocal (2017) for a more recent overview.

ends up convicting the guilty by basing every verdict on a series of coin flips. Intuitively, the correctness of the verdicts does not guarantee fairness. Legal fact-finders are not concerned with mere accuracy: plausibly, they also aim to instil public confidence in the trial as an institution. They must ensure not only that the verdicts are right and just, but also that they are *perceived* to be right and just. To do so, an anti-luck condition is needed: just like knowledge is incompatible with luck, so is the evidence upon which a legal verdict is based.

In light of these common features, striking similarities between the project of analysing knowledge and the attempt to provide an epistemic account of individualised evidence are brought into focus. According to Goldman (1967), knowledge requires causal relations, and so does individualised evidence according to Judith Thomson (1986). Nozick (1981) argued that knowledge requires sensitivity to truth, and so does individualised evidence according to Enoch, Fisher and Spectre (2012). Sosa (1999), Pritchard (2005, 2012a) and Kelp (2013) have proposed a safety condition on knowledge: other legal epistemologists – Pritchard included – have followed suit and imposed a safety condition on individualised evidence (Pritchard 2018b, Pardo 2018).

I now want to pause to register the following: the two debates significantly mirror each other. Within traditional epistemology, these conditions were proposed to fix the tripartite analysis of knowledge (allegedly) refuted by Gettier.<sup>17</sup> Importantly, an adequate solution to the Gettier problem presupposes a successful anti-luck condition: after all, the tripartite analysis clearly lacked one, as Dancy (1985: 134) and Pritchard (2015: 94) have later observed. Within legal epistemology, the same conditions were proposed to identify individualised evidence, which, like knowledge, also presupposes a similar non-accidental link to truth. This is the structural connection between knowledge and individualised evidence, and legal epistemologists have taken notice of it: more or less explicitly, they have assumed that the very same anti-luck condition on knowledge will also be necessary for (and so properly characterise) individualised evidence. Redmayne (2008: §4) dubs such an approach “knowledge-based” approach.

The knowledge-based approach enjoys initial plausibility, but I worry that it is not being pursued correctly (perhaps it’s not even worth pursuing; more on this in §5). Let’s pay heed to a key neglected detail. It bears noting that the knowledge-based approach proceeds in a somewhat *indirect* fashion: rather than on knowledge, so far the action has been on a *separate* anti-luck condition on knowledge. Building on this, I want to highlight an important and yet hitherto overlooked feature of the knowledge-based approach. The structural connection between knowledge and individualised evidence has been understood in a very *indirect* fashion: in fact, the focus is on separate anti-luck conditions on knowledge like (appropriate) causation, sensitivity and safety. The knowledge-based approach assigns explanatory priority to these separate causal or modal anti-luck conditions on knowledge and, as such, it turns out to be an essentially *indirect* approach.

The indirect way in which legal epistemologists have developed the knowledge-based

---

<sup>17</sup>See Ichikawa and Steup (2018), and Antognazza (2015) for a more accurate historical reconstruction.

approach deserves special scrutiny. The assumptions underlying it are suspicious, and rest on shaky foundations. These separate anti-luck conditions on knowledge featured in once prominent analyses, but, after all, the project of analysing knowledge has not borne much fruit. As Williamson (2000: 2-5, 30) observes, given the endless plethora of counterexamples faced by every purported analysis of knowledge, it is reasonable to conclude that none of them is problem-free. This invites the following question: granting that there is a structural connection between knowledge and individualised evidence, why think that such connection is best couched in terms of problematic anti-luck conditions on knowledge? Once imposed on individualised evidence, these conditions are likely to face the same difficulties that they face with knowledge. And this is precisely what we find. If legal epistemologists aim to uncover the structural connection between knowledge and individualised evidence, they'd better learn from the problems of anti-luck conditions on knowledge in lieu of simply reimposing them on individualised evidence.

In the remainder of the chapter, I will substantiate this thought and show that, once imposed on individualised evidence, prominent anti-luck conditions on knowledge run into the same familiar and perhaps unavoidable difficulties that we see in post-Gettier epistemology. My strategy paves the way for a potentially far-reaching conclusion, which I revisit more extensively in §5: if I am right, no single causal or modal anti-luck conditions on knowledge can succeed as the right anti-luck condition on individualised evidence. As I will clarify, this conclusion has wider repercussions not only for the relevance of traditional epistemology to the understanding of individualised evidence, but also for legal epistemology more in general.

Having clarified these issues, I now proceed to assess three prominent and indirect knowledge-based approaches, starting with Thomson's causal account of individualised evidence.

## 7.5 Causation

Two elements stand out in **Accident** and **Prisoners**. Firstly, statistical evidence lacks the right type of *direct connection* with the relevant facts. Secondly, statistical evidence does not eliminate the possibility of luck: impositions of liability and guilt based on statistical evidence can still be correct as a matter of luck. Judith Thomson (1986) takes these observations as starting points, and develops an account of individualised evidence in terms of *causal relations*.

Her account comprises two theses. First, individualised evidence plays an essential role: it provides a non-accidental *guarantee* of guilt and liability. Second, only causally specified evidence provides such guarantee. Putting the two together, Thomson concludes that individualised evidence is causally specified: causal relations provide a guarantee which is meant to be immune to malign luck. Crucially, the resemblance with Goldman's (1967) causal analysis of knowledge is stark:

**Causal Analysis of Knowledge.** S knows that *p* if and only if S's belief that *p* stands

in an appropriate causal relation to the fact that makes  $p$  true.

**Causal Individualised Evidence.** Individualised evidence guarantees that  $p$  if and only if it stands in an appropriate causal relation to the fact that makes  $p$  true.

In keeping with the indirect knowledge-based approach, Thomson moves from an anti-luck causal condition on knowledge to an anti-luck causal condition on individualised evidence. Her account is motivated by the overarching role that luck plays in making a verdict of guilt based on statistical evidence only accidentally correct (Thomson 1986: 214), so I'd better be clear on how luck relates to statistical and individualised evidence.<sup>18</sup>

Following Pritchard (2012a), I read Thomson as holding that individualised evidence, unlike statistical evidence, should be incompatible with *veritic* luck, a type of luck which makes a belief only accidentally true and thus prevents it from constituting knowledge. Veritic luck operates at different levels. In the well-known barn façade example originally credited to Gettier, veritic *environmental* luck makes a belief luckily true for reasons having to do with the external environment: the belief is true but it could easily have been false. In standard Gettier cases (for instance, the sheep in the field case discussed in Chisholm 1966: 105), veritic *intervening* luck interferes in a more direct manner: an otherwise false belief accidentally hits the truth only courtesy of an intervention of luck. In what follows, I assume that both types of veritic luck are incompatible with individualised evidence (perhaps illegitimately: I return to this below).

We now get a better purchase on Thomson's causal account: while statistical evidence is compatible with veritic luck, causal individualised evidence is importantly immune to it. However, her proposal is plagued by the same problems that beset a causal theory of knowledge, and one of such problems has to do precisely with veritic luck. *Pace* Thomson, causal relations are compatible with environmental luck, as we learned from Goldman's early causal theory of knowledge. Consider the following case:

**Fake Knives.** The famous detective Sherlock Holmes and his loyal sidekick John Watson are trying to solve a murder case. The manner of death of the victim is fairly standard (stabbed to death multiple times with a knife), but the crime scene is peculiar: it's filled with countless *seemingly identical* bloody knives. In order to throw off the detectives, the murderer made sure not only to cover each knife with the blood of the victim, but also with the fingerprints of random innocent people. Crucially, only one of them was used as the actual murder weapon – it's been used to kill the victim and it bears the fingerprints of the murderer. Sherlock inspects the scene, and by sheer luck he picks the very knife which was used as the murder weapon. Watson brings the knife as hard evidence to trial, and the murderer is convicted on the basis of the incriminating bloody knife.

This is a fake-barn case directed against Thomson. Its correct description is the following. The evidence is individualised and causally specified, and appropriate causal relations ob-

---

<sup>18</sup>We should grant to Thomson that causation is not probabilistic: if paired with more recent probabilistic theories of causation (see Hitchcock 2021: §2), her account of individualised evidence would look very different.

tain between the relevant facts. The bloody knife is present on the crime scene *because* it was used as the murder weapon; the bloody knife bears the fingerprints of the murderer *because* they used it to commit the crime; the murderer is guilty *because* they used the knife to commit the crime and *because* the knife bears their fingerprints; and so on. However, just like the risk of error is too high to secure knowledge in fake-barn county, so the risk of error is too high in the case of a conviction based on this piece of causally specified individualised evidence. The verdict is only luckily correct. It could have easily been mistaken, and based on a seemingly identical knife which would have further led to the wrongful conviction of an innocent. Clearly, the verdict is correct only courtesy of a double stroke of luck: one bad (there are many identical murder weapons), one good (Sherlock happens to pick the right one). Causal individualised evidence is compatible with undermining environmental luck, and thus fails to provide a non-accidental guarantee.

I offered this case as a proof of concept that we can still have causally specified individualised evidence and malign veritic luck.<sup>19</sup> In order to better appreciate this point and to abstract away from too complicated thought experiments, I would like to offer two Gettier-style variations of stylised courts cases like **Accident** and **Prisoners**:

**Fake Blue Buses.** While driving back home, a vehicle suddenly hits Jane's car. In the accident, she breaks her leg. Thankfully, a passer-by witnesses the event: it was a blue bus that hit Jane's car, and so does the witness reports. However, there's a twist to the story: many more buses painted in blue but actually driven by the competing red bus company drivers were also present on the scene of the accident and also operated a route at the same time of the event. The eyewitness correctly identifies the only genuine and liable blue bus, but they do so only as a result of sheer luck.

**Twin Prisoners.** 100 prisoners are exercising in the prison yard. Suddenly 99 of them attack the guard, putting into action a plan that the 100th prisoner knew nothing about. The 100th prisoner played no role in the assault and could have done nothing to stop it. Thankfully, a guard witnesses the event, and correctly identifies prisoner number 58 as guilty. However, there's a twist to the story: prisoner 58 has an identical twin brother who happens to be exactly the 100th innocent prisoner. The eyewitness correctly identifies the guilty prisoner, but they do so as a result of sheer luck.

To repeat: these are further fake-barn variations of **Accident** and **Prisoners**. Here too, the evidence is individualised and causally specified, yet it fails to exclude veritic luck due to misleading features of the scene. Again, appropriate causal relations obtain between the relevant facts. The blue bus company is liable because a blue bus hit Jane's car, and the passer-by saw a blue bus hitting Jane's car because a blue bus did in fact hit Jane's car. Prisoner 58 is guilty because they assaulted the guard, and the eyewitness saw prisoner 58 assaulting the guard because prisoner 58 did in fact assault the guard. However, even

---

<sup>19</sup>You may feel discomfort at advancing the debate on the basis of such seemingly contrived cases. I do too: I return on this in §5 and point out this fundamental methodological problem for the knowledge-based approach.

if based on causally specified individualised evidence, the verdict is only luckily correct: the passer-by could have too easily looked at a fake blue bus driven by a red bus driver, and then the piece of individualised evidence would have incorrectly determined that the blue bus company is liable. Similarly, the eye-witness could have too easily mistaken the guilty prisoner 58 for their innocent identical brother, and then the evidence would have incorrectly determined that prisoner 100 is guilty. The point remains: causally specified individualised evidence is still compatible with an only (veritically) luckily correct verdict.

We have now come full circle. Thomson argues against statistical evidence because verdicts based on it can be correct as a matter of luck, and she goes on to impose a causal anti-luck condition on individualised evidence by leaning on a causal theory of knowledge. However, her proposal suffers from the same problem she raises, since causal relations are equally compatible with veritic environmental luck.<sup>20</sup> Tellingly for my purposes, my objection against her causal account mirrors the counterexample against the causal analysis of knowledge.

## 7.6 Modal Conditions

The same problem affecting a causal theory of knowledge also spells trouble for a causal account of individualised evidence. Causal relations are compatible with veritic luck, and therefore fail as anti-luck condition on both knowledge and individualised evidence. One way of diagnosing this problem is to hold that individualised evidence should appropriately *track* the truth; another is to maintain that it should not *too easily* support a falsehood. These observations pave the way for two modal anti-luck conditions on knowledge, sensitivity and safety. Can they do better than causal relations as anti-luck conditions on individualised evidence? In the following two sections, I answer negatively. In §4.1, I show that sensitive evidence also falls prey to veritic luck, and point out further difficulties with other canonical instances of individualised evidence. Next, in §4.2., I argue as follows. Even granting that safety is a successful anti-luck condition, the principle does not carve at the right epistemological joints: in fact, it fails to distinguish between statistical and individualised evidence.

### 7.6.1 Sensitivity

The sensitivity condition on knowledge was introduced by Robert Nozick, who suggests that if S knows that  $p$ , then S's belief that  $p$  tracks the truth. Following Nozick (1981),

---

<sup>20</sup>There is increasing tendency to accept that propositional knowledge is compatible with environmental luck, most notably in Sosa's bi-level virtue epistemology (Sosa 2015: 77-81). I remain neutral on whether propositional knowledge is compatible with environmental luck: it should be granted that agents run a high possibility of error in these scenarios, and that causally specified evidence does not insure against environmental luck. This suffices to raise the objection against causally specified evidence, independently of the potentially more controversial claim that knowledge and environmental luck are incompatible (e.g., Pritchard 2012a: §3). Notice further that causally specified evidence does not insure against intervening luck either: I could equally press a case of deviant causal chains against Thomson, but here I've decided to focus on environmental luck to highlight the analogy with Goldman's causal theory of knowledge.

I break down the sensitivity principle into two twin conditions, variation and adherence. Here is a full-blown statement of sensitivity:

**Sensitivity.** S knows that  $p$  if and only if: had  $p$  been false, then S would not have believed  $p$  (variation); and, had  $p$  been true, then S would have believed  $p$  (adherence).

Enoch, Fisher and Spectre (2012) build on Nozick's analysis of knowledge and impose a sensitivity condition on individualised evidence.<sup>21</sup> Accordingly:

**Sensitive Evidence.** Individualised evidence  $e$  is sensitive to  $p$  if and only if: had  $p$  been false, then S would not have believed  $p$  on the basis of  $e$  (variation); and, had  $p$  been true, then S would have believed  $p$  on the basis of  $e$  (adherence).

With an eye on the problems encountered by a sensitivity condition on knowledge, I shall develop two main foci of attack against sensitive evidence, which is either too weak or too strong. Too weak: it fails to safeguard against veritic luck. Too strong: it rules out insensitive but acceptable instances of individualised evidence. Let me take each point in turn.

First off, sensitivity does not preclude (Gettier-style) intervening luck. In fairness to sensitivity theorists, their principle wasn't originally offered as an explicit anti-luck condition. However, several authors - Nozick included - have conceded that sensitivity should preclude standard Gettier cases, which in turn involve intervening veritic luck.<sup>22</sup> Sensitivity does not meet this desideratum: more ingenious Gettier cases can be raised with equal force also against sensitivity.<sup>23</sup> Here is a particularly apt one:

**Lucky Cold Hit.** A team of detectives is investigating a murder. The evidence is scarce, but some biological traces are available. Genetic material is found at the crime-scene: it's a strand of hair belonging to the perpetrator. It gets tested, and the results bring in a new suspect, Mr. X, who has indeed committed the crime. Upon further investigation, Mr. X is found guilty and convicted on the basis of the DNA evidence. There is, however, a twist to the story. The DNA found on the scene matched the new suspect only by luck: courtesy of a previous lab error, Mr. X's long deceased monozygotic twin brother, Mr. Y, was accidentally registered in the DNA database under the name of Mr. X. While it's easy to see why the mistake was made (their DNA profile is identical), the lab error is nevertheless the chief reason why the DNA database included Mr. X's profile in the first place.

---

<sup>21</sup>To forestall misunderstanding, I flag that sensitivity plays only a *partial* role in Enoch and colleagues' overall account of individualised evidence. While they assume a sensitivity account of individualised evidence, to avoid what they call knowledge-fetishism their incentive-based explanation carries more theoretical weight than sensitivity (Enoch et al. 2012: 220-223). But sensitivity remains relevant to their incentive-based explanation: they acknowledge that "the epistemic story and the incentive-based story are closely knit and interestingly related in light of their similar structures and ramifications in the legal arena" (Enoch and Fisher 2015: 565).

<sup>22</sup>Pritchard (2005: 158; 2013: §2) discusses the role of sensitivity in handling Gettier cases. Nozick (1981: 173) admits that his sensitivity condition should serve as anti-Gettier condition.

<sup>23</sup>See also Goldberg's **Movie Set** (Goldberg 2015: 279), and Becker's **Vase Hologram** (Becker 2018: 124).

The case I offer is a Gettierised cold-hit DNA match. As Smith (2018: 1215) also notes, DNA evidence is usually sensitive, and satisfies sensitivity's two-fold condition. In **Lucky Cold Hit**, the evidence satisfies variation. Had Mr. X not committed the murder, he wouldn't have left any genetic material at the crime scene and there would have been no hit. On such basis, the investigators *would not* have believed him to be guilty. Likewise, the evidence satisfies also adherence. Had Mr. X committed the murder in relevantly similar circumstances, he would have left genetic material at the crime scene, and the hit would have taken place. On such basis, the investigators *would have* still believed him to be guilty. Yet in this case sensitive DNA evidence tracks the truth and supports a correct verdict only courtesy of intervening luck: the DNA belongs to the culprit, but the hit is lucky and the match accidental. In this respect, sensitivity is too weak to provide a successful anti-luck condition.

In another respect however, sensitivity is too stringent: if knowledge required sensitivity, we'd have too little of it. Similarly, once sensitivity is imposed as necessary condition, the resulting conception of individualised evidence is extremely impoverished. Sensitivity theorists struggle to account for cases of safe but insensitive inductive knowledge, as Sosa (1999: 145) first noticed. In like manner, pieces of inductive evidence are safe but equally insensitive. Consider:

**Kantian Eye-Witness.** Maria likes to take strolls. She does this with extreme reliability, and her habit is often compared to Kant's famous afternoon walks. During her strolls, Maria regularly walks on the street where the accident takes place, and regularly sees a blue bus operating the route. The day comes, and a blue bus hits the car. Maria doesn't witness the event, but testifies that she spotted a blue bus shortly before the time of the crash, and that she has seen a blue bus operating the route in that very street every other day until that day. To back up her statement, she lists every time she's taken her walk and seen a blue bus operating the route around the time and place of the crash. The list is long, and provides (inductive) reason to conclude that a blue bus is liable.

This is a case of inductive eye-witness testimony (and, *a fortiori*, individualised evidence). Sensitivity gets the wrong result here: Maria's testimony may be insensitive, but it's nevertheless acceptable in virtue of being safe. Her testimony is insensitive: in the closest possible world where a blue bus is not liable (perhaps because a red bus mysteriously went off-track and hit the car), one forms a false belief on the basis of her testimony. Still, her testimony is safe, and thus acceptable: in more relevant close worlds, a blue bus is liable, and one *does* form a true belief on the basis of her testimony. Because of sensitivity, this key aspect simply goes overlooked.

To further appreciate, on more general grounds, why sensitivity is a too strong necessary condition for individualised evidence there's one final case I want to offer - an epistemic Frankfurt case. Notably, such cases have been successfully deployed against safety conditions (Kelp 2016), but they apply with equal (if not major) force also against sensitivity. Consider:

**Scheming Prisoner.** 100 prisoners are exercising in the prison yard. Suddenly 99 of them attack the guard, putting into action a plan that the 100th prisoner knew nothing about. Prisoner 54 is guilty, but another scheming prisoner is not aware of this and wants to make sure that prisoner 54 is punished anyway due to long-standing resentment against him. The scheming prisoner hatches his plan to frame the assault on prisoner 54, fully ready to tamper with any piece of exculpatory evidence. However, he finds an authentic videotape that already implicates prisoner 54 and, happy with the result, he doesn't intervene.

Intuitively, the videotape provides insensitive but acceptable individualised evidence of guilt. The evidence is insensitive: in the closest possible world where prisoner 54 is innocent, the scheming prisoner intervenes, tampers with the evidence (the videotape) and frames him. Had  $p$  been false, one would have believed  $p$  anyway given the evidence. More precisely: had prisoner 54 been innocent, one would have believed him to be guilty anyway thanks to the suitable intervention of the scheming prisoner. Yet, even if insensitive, the evidence is acceptable. Any reason to discard it seems unmotivated: unlike standard Gettier and fake-barn cases, there's no *actual* environmental or intervening luck involved. There's no deviant causal chain, and no clear indication of any salient impropriety (epistemic or otherwise). Because of sensitivity, we are forced to deny the intuitive status of acceptable individualised evidence to the videotape. The lesson to learn here is general: sensitivity is too strong a condition for individualised evidence.<sup>24</sup>

Let me take stock. I've developed a two-pronged attack against sensitive evidence. Firstly, sensitivity is too weak to safeguard against intervening luck: Gettier-style cases are easy to find and generate (**Lucky Cold Hit**). Secondly, sensitivity is also too strong: it rules out safe but insensitive inductive evidence (**Kantian Eye-Witness**), and delivers the wrong verdict on plain cases of acceptable but insensitive individualised evidence (**Scheming Prisoner**). Here too, problems faced by a sensitivity condition on knowledge apply also to sensitive evidence.

## 7.6.2 Safety

Neither causation nor sensitivity succeed in excluding veritic luck. One final candidate left to consider is another anti-luck condition proffered by epistemologists, the safety principle. Pritchard also moves from a safety condition on knowledge to one on individualised evidence:

The modal condition that we are imposing on legal evidence in this regard is what is known as a safety condition. Such a condition is generally regarded as being a necessary condition for knowledge, and also – relatedly – being the condition that excludes

---

<sup>24</sup>One may worry that epistemic Frankfurt cases put pressure on *any* modal anti-luck condition on knowledge and individualised evidence (both sensitivity and safety). After all, such cases were originally designed to show that safety – rather than sensitivity – is a too strong condition (Kelp 2016: §2.1). The reader is free to draw this stronger conclusion, but I aim for less. I offer the Frankfurt case to show that sensitivity is too strong, and I prefer to remain non-committal on different (and perhaps less clear-cut) verdicts pertaining to safety. As I will show in the next subsection, safety accounts of individualised evidence are problematic on independent grounds.

the kind of epistemic luck/risk which is incompatible with knowledge. (Pritchard 2018b: 117)

First, it bears noting that several authors have questioned the success of safety as anti-luck condition: a true belief can be lucky even if formed through a safe method.<sup>25</sup> Thus, *pace* Pritchard, safety may not succeed to exclude the kind of epistemic luck which is incompatible with knowledge. However, the problem is worse than this: even if safety were a successful anti-luck condition, it would not draw the line in the right place. In fact, in both its weak and strong formulations, safety does not distinguish between statistical and individualised evidence.

Pritchard argues that while statistical evidence is unsafe, individualised evidence is safe (Pritchard 2018b: §3). This is because, on the basis of the former, the modal risk of forming a false belief is high: little would have to change in the actual world for a belief based on statistical evidence to be false. Statistical evidence is *unsafe*: in similar nearby worlds, beliefs based on it are false. Conversely, on the basis of the latter, the modal risk of forming a false belief is low: a lot would have to change in the actual world for a belief based on individualised evidence (e.g., eye-witness testimony) to be false. Individualised evidence is *safe*: in similar nearby worlds, beliefs based on it continue to be true. These are the core claims of Pritchard's safety account.

Pritchard does not commit himself to a specific formulation of safety. Thus, the success of his safety account turns on a preliminary terminological disambiguation and two distinct questions. First, we need to get clear on whether safety is understood weakly or strongly. Having clarified this, we can properly assess the safety account by focussing on the following two questions. Given a more refined formulation of the safety principle, is Pritchard right in claiming that statistical evidence is unsafe (Question 1) and that individualised evidence is safe (Question 2)? Let's begin with the weaker formulation of safety:

**Weak Safety.** S believes *p* on the basis of safe evidence if and only if S's belief that *p* is true in the actual world and in *most* close possible worlds in which S forms the belief that *p* on the basis of the same evidence S uses in the actual world, *p* is true.<sup>26</sup> (Greco 2012b: 196)

Statistical evidence satisfies weak safety. Recall our initial examples: let's grant that prisoner 54 is guilty and that a blue bus hit Jane's car. These are not random aleatory events: as Gardiner (2020: §6) notices, guilty prisoners may riot also in close worlds. Ditto

---

<sup>25</sup>See for instance Hiller and Neta (2007: 208); Lackey (2006: 288); Coffman (2010: 246); Goldberg (2015: 275). To appreciate the strength of such counterexamples, it's worth emphasising that while safety is offered as a merely necessary condition on knowledge, it is nevertheless meant to be sufficient to exclude knowledge-undermining luck (Pritchard 2013: 158).

<sup>26</sup>While standard construals of safety assume truth, there's an additional reason to stipulate that *p* is true: the truth of *p* impacts on its modal profile. As Ebert et al. observe (2018: 10-11), no world is more similar to the actual world than the actual world itself. It's thus difficult to make assessments on safety without already taking a stand on whether the belief in question is true or false or on whether the defendant is guilty or innocent. This is a structural feature of Pritchard's conception of safety and risk, as Smith (2018: 1204-1205) also notices.

for liable bus drivers, who may equally drive negligently in close worlds too. Statistical evidence can be weakly safe: there can be most close worlds where prisoner 54 continues to assault the guard, and a blue bus still hits Jane's car. Thus, on its basis, one does form a true belief in most close worlds. The vignettes are admittedly underdescribed, but nothing in their description suggests that the modal profiles of events like a riot or an accident are fragile: plausibly, prisoner 54 does not just riot on a whim, nor does the bus driver "just so happen" to hit Jane's car. Once these events obtain in the actual world, they also plausibly obtain in most close worlds.<sup>27</sup>

The foregoing points provide an answer to Question 1. In the cases of interest, beliefs based on sufficiently strong statistical evidence are weakly safe: contrary to Pritchard's account, in nearby worlds beliefs based on it continue to be true. The prospects for the safety account may start to look dim by now, but not all is lost for the safety theorists, who can still appeal to a stronger formulation of safety:

**Strong Safety.** S believes *p* on the basis of safe evidence if and only if S's belief that *p* is true in the actual world and in *every* close possible world in which S forms the belief that *p* on the basis of the same evidence S uses in the actual world, *p* is true.

(Greco 2012b: 196)

Strong safety is a more promising principle for ruling out statistical evidence, which fails to satisfy this stronger requirement. After all, there will be one or more close possible worlds where, on the basis of statistical evidence, one will falsely believe that prisoner 54 attacked the guard, or that a blue bus hit Jane's car. Accordingly, strong safety gives a compelling negative answer to Question 1. So far, so good. What about Question 2? Here, safety theorists have additional work to do: they have to show not only that statistical evidence is not (strongly) safe, but also that individualised evidence is (strongly) safe. However, under a strong formulation of the safety principle this is not feasible: just like statistical evidence, individualised evidence also fails to satisfy a stronger safety requirement.

Here is why. Take a canonical piece of individualised evidence such as eye-witness testimony. Inevitably, there will be a close possible world where one forms a false belief on its basis. All we need is a close enough world where a very improbable possibility is actual. Consider:

**Dishonest Eye-witness.** While driving back home, a vehicle suddenly hits Jane's car. She realises that the vehicle was a bus, but she cannot discern the colour. A huge crowd of bystanders witness the event. Jane asks to one of them the colour of the bus which hit the car. The bystander answers sincerely: it was a blue bus.

Every other bystander would have answered in the same way – all except one. In

---

<sup>27</sup>Safety theorists can respond to Gardiner's challenge in (at least) three ways. First, they may strengthen the safety principle and deny that statistical evidence is safe. However, as I explain below, this is a costly answer. Second, they may index the safety principle to specific belief-formation methods. This is also a costly answer, likely to raise the generality problem famously pressed by Conee and Feldman (1998). Third, they may restrict the focus to normal worlds instead of similar close worlds (see Smith 2016: Chapter 6 for a comparative discussion of safety and normalcy). I take this to be the best response, but I lack the space to consider it fully here.

fact, unbeknownst to Jane, among the bystanders there's also a dishonest eye-witness, who works for the blue bus company and is prepared to lie to cover for his company. Luckily, Jane picks one of the many sincere bystanders instead of the only dishonest eye-witness.<sup>28</sup>

This is our close but improbable possibility. Jane could easily have formed a false belief on the same basis: there's a very similar close possible world where she casually picks the dishonest eye-witness instead of the sincere bystander. Since strong safety does not tolerate the possibility of error in *any* close world, this canonical instance of individualised evidence is not strongly safe. Again, contrary to Pritchard's account, in nearby worlds beliefs based on it are false.

Overall, strong safety places too demanding a requirement on individualised evidence, which, plausibly enough, supports the truth of a proposition in *most* close worlds rather than in *every* close world. As a necessary condition, strong safety comes at a high cost: while it may rule out statistical evidence, it excludes also canonical pieces of individualised evidence.

At this point, each of Pritchard's core claims is cast into doubt. Since it satisfies weak safety, it's not the case that statistical evidence is unsafe. And, since it does not satisfy strong safety, it's not the case that individualised evidence is safe. Safety theorists now face a dilemma: either individualised evidence is weakly safe [*Option 1*], or it is strongly safe [*Option 2*].<sup>29</sup> Suppose that they choose [*Option 1*]: if individualised evidence is weakly safe, then so is statistical evidence. Weak safety fails to distinguish between the two. Suppose that they choose [*Option 2*] instead: if individualised evidence is *strongly* safe, then neither statistical evidence nor canonical instances of individualised evidence (e.g., eye-witness testimony) are strongly safe. Strong safety also fails to distinguish between the two.

This is a minimum desideratum for a successful safety-based account: individualised evidence should have a modal epistemic quality (weak or strong safety) that statistical evidence lacks.<sup>30</sup> Safety does not deliver on this front. No matter which option is picked, safety does not carve at the relevant epistemological joints: it does not distinguish between statistical and individualised evidence. It looks like the right anti-luck condition on individualised evidence just can't be found; I now proceed to say more on why it can't be found.

---

<sup>28</sup>To keep the discussion manageable, I ask the reader to bracket complicated issues pertaining to further defeating and corroborating evidence Jane might acquire during later stages of her inquiry.

<sup>29</sup>Safety-theoretic accounts of knowledge face similar objections (e.g., Greco 2007: §1; Sosa 2015: 119; Gardiner 2020: §6). The dilemma I raise here is a hybrid and hence novel one: in [*Option 1*], I borrow elements from Gardiner's objection. In [*Option 2*], I draw on Greco's and Sosa's versions.

<sup>30</sup>Blome-Tillman (2015) also implicitly accepts this desideratum. He objects to the sensitivity and causal accounts of individualised evidence by showing that statistical evidence can also be sensitive or causally specified.

## 7.7 Diagnosis

At this point, let's focus on the bigger picture: no account of individualised evidence in terms of single causal or modal anti-luck conditions on knowledge is or can be correct. Given the failure of these conditions on knowledge, we should expect them to fail also when imposed on individualised evidence. The expectation is now confirmed: these conditions fail on *both* knowledge and individualised evidence. But what should we make of this double failure? I suggest to make an inference to the best explanation: just like any analysis of knowledge in terms of separate causal or modal anti-luck conditions is unsuccessful, so will be any account of individualised evidence in terms of the same anti-luck conditions. Here's my diagnosis: no single causal or modal anti-luck condition on knowledge can succeed as anti-luck condition on individualised evidence. This also calls into question the relevance of these anti-luck conditions to the understanding of individualised evidence and to legal epistemology in general.

Before looking at the broader consequences of this diagnosis for legal epistemology, it's worth dwelling on its independent plausibility. That veritic luck is hard to exclude is a familiar take-home lesson from the post-Gettier literature: purported causal and modal conditions on knowledge fail to capture the right non-accidental connection to truth, nor insure against veritic luck. The problem is general: as Zagzebski (1994: 72) explains, veritic luck strikes "as long as there is a small degree of independence between truth and the other conditions on knowledge".

Given the inescapability of veritic luck, single causal or modal anti-luck conditions on knowledge are intrinsically defective. This problem is also general. Here's a telling quote from Goldberg:

No purely modal condition (of the sort epistemologists have used to analyze knowledge) will suffice as the anti-luck condition on knowledge, since we can concoct examples in which it is simply a matter of luck that the method has that modal property. (Goldberg 2015: 279)

So much for knowledge. However, and crucially for my purposes, this problem runs deeper. In fact, as explained in §3, the knowledge-based approach championed by some legal epistemologists takes knowledge and individualised evidence to be structurally connected by a similar anti-luck condition: Goldberg's pessimistic remarks apply also to individualised evidence. For sake of vividness, let me make this last point from a different angle. Consider this rough and ready equation, where  $x$  denotes a causal or modal anti-luck condition distinct from knowledge:

$$\text{Knowledge} \equiv \text{Truth} + \text{Belief} + x$$

These days, this equation has fallen out of favour. We learn from post-Gettier epistemology that while truth and some separate causal or modal anti-luck condition may be individually necessary, they're nevertheless jointly insufficient to exclude veritic luck and obtain knowledge. Despite many attempts, no uncontroversial  $x$  has been found. Suppose that,

drawing on a strong inductive basis, we throw in the towel and follow Williamson (2000: 3-4) in thinking that no  $x$  can be found.<sup>31</sup> Then the same inductive basis gives us reason to reject also this rough and ready equation, where  $x$  still denotes an anti-luck condition distinct from knowledge:

$$\text{Individualised Evidence} \equiv \text{Truth} + x$$

Proponents of the knowledge-based approach accept or assume some version of this equation.<sup>32</sup> In this chapter, I have considered three prominent candidates for  $x$ : a causal and two modal anti-luck conditions. However, just like these conditions fail to identify the non-accidental connection to truth which is necessary for knowledge, likewise they fail to identify the non-accidental connection to truth which is taken to be necessary for individualised evidence. What is more, they fail for similar reasons: objections to separate causal and modal anti-luck conditions on knowledge apply also to accounts of individualised evidence in terms of the same anti-luck conditions. Crucially, these objections are inescapable: they generalise to any account of individualised evidence built around a causal or modal anti-luck condition distinct from knowledge itself.

These points are significant for legal epistemology. The failure of single causal or modal anti-luck conditions on both knowledge and individualised evidence suggests that the notion of individualised evidence can't be successfully elucidated in terms of these single causal or modal anti-luck conditions. Because it employs these defective anti-luck conditions, the knowledge-based approach is unfruitful and fundamentally limited. Firstly, such approach consists in imposing on individualised evidence the very same anti-luck conditions that already fail when imposed on knowledge: as such, this way of proceeding appears to be a non-starter. Secondly, because it uses rather contrived counterexamples made popular in mainstream epistemology the knowledge-based approach fails to engage with the subtle and complicated realities of existing legal systems. The approach assumes a somewhat omniscient observer (standardly absent in legal courts), and, in the case of modal conditions, a somewhat precise and privileged access to the closeness of possible worlds that determine judgements on safety and sensitivity (a privileged access which is also standardly absent in legal courts). Thirdly, no matter which specific anti-luck condition is picked, anti-luck conditions in general fail to provide a meaningful distinction between statistical and individualised evidence. As we have seen, they do not draw the line in the right place: that there is no real distinction between statistical and individualised evidence remains a live option which the knowledge-based approach can't rule out. Accordingly, despite being motivated by what appears to be a common anti-luck condi-

---

<sup>31</sup> See Cassam (2009: 21-22) and Williamson (2009: 290) for discussion on the unanalysability of knowledge.

<sup>32</sup> The equation is schematic, and does not represent (or attribute) a reductive analysis. I am concerned with anti-luck conditions on individualised evidence: while these conditions are meant to be sufficient to exclude luck, they are not meant to be sufficient for a positive verdict. Similarly, as noted above, anti-luck conditions on knowledge are meant to be sufficient to exclude knowledge-undermining luck, not sufficient for knowledge. I remain neutral on whether individualised evidence demands more than a successful anti-luck condition to suffice for a positive verdict. Even a fully successful anti-luck condition may be a necessary but still insufficient condition for this more demanding role.

tion on knowledge and individualised evidence, the knowledge-based approach is not very fruitful for legal epistemology. Given the focus on too many artificial cases, it is ill-suited to advance the understanding of legal evidence. And since it fails to provide a successful distinction between statistical and individualised evidence, it is also ultimately inconclusive. Taken together, these considerations suggest that as far as veritic luck is concerned, epistemology and the law seem to come apart.

This last point is best appreciated in the context of other important objections against the relevance of epistemology to the understanding of evidence law. The problems that single anti-luck conditions on knowledge incur when extended to individualised evidence speak in favour of Allen's complaint against a naive and simple application of epistemic concepts to the law. As he puts it (Allen 2021a: 255), this application often results in relying on "weird hypotheticals" that place "impossible epistemological demands" on a crucial type of legal evidence like individualised evidence.<sup>33</sup> The problems of the knowledge-based approach also bolster Fratantonio's recent case against a purely epistemic solution to the puzzle of statistical evidence (Fratantonio 2021). Borrowing from her useful taxonomy (Fratantonio 2021: §2), if legal epistemology is conceived as a descriptive project that aims at elucidating legal concepts (e.g., individualised evidence) in epistemic terms (e.g., causal or modal anti-luck conditions), then the main anti-luck conditions on knowledge fail to achieve this clarificatory aim: after all, they don't successfully distinguish between statistical and individualised evidence. And if legal epistemology is conceived as a normative project concerned with why an ideal legal system *should not* use statistical evidence, then the main anti-luck conditions on knowledge also fail to achieve this aim for similar reasons (we are still left without a meaningful distinction between statistical and individualised evidence). Overall, anti-luck conditions seem irrelevant for the pursuit of both a descriptive and a normative project in legal epistemology.

These remarks align well with further compelling attempts that build on normative considerations to cleave epistemology from evidence law. Since the knowledge-based approach has proven unsuccessful in tracking any substantial normative difference between statistical and individualised evidence, Enoch, Fisher and Spectre may be right in holding that epistemology is not "intrinsically and directly relevant to normative evidence law theory" (Enoch, Fisher and Spectre 2021: 85). Similarly, the limits of the knowledge-based approach provide further reason to think, as Ross (2021a) is keen to emphasise, that courts of law are subject to normative requirements fundamentally distinct from those applying to individual subjects routinely discussed in epistemology. There seems to be room for fruitful interactions between epistemology and theory of evidence law, but neither the knowledge-based approach nor the focus on anti-luck conditions are instances of such interactions.<sup>34</sup>

I would like to note a few important differences with these influential objections to the relevance of epistemology to the understanding of legal evidence. Because it is circum-

---

<sup>33</sup>See Allen (2021a, 2021b, 2021c) and Allen and Leiter (2001) for an illuminating elaboration of this criticism.

<sup>34</sup>Pundik (2011) also agrees.

scribed to anti-luck conditions, the point made here is more modest: unlike Fratantonio's objection, the arguments in this chapter do not target epistemic solutions to the puzzle of statistical evidence that do away with anti-luck conditions on knowledge. Similarly, unlike Enoch, Fisher and Spectre's criticism, the arguments offered here don't generalise to every epistemic notion. However, these arguments do retain the spirit of these influential points against the relevance of epistemology for legal evidence and also contribute to weaken such relevance.<sup>35</sup>

Having clarified this, I can take stock and conclude. In light of an intuitively similar anti-luck condition, one approach in legal epistemology has consisted in moving from an anti-luck condition on knowledge to one on individualised evidence. This *modus operandi* has been found wanting: no causal or modal anti-luck condition on knowledge can succeed as anti-luck condition on individualised evidence. The chief moral to draw from post-Gettier epistemology is that causal and modal conditions on knowledge fail as satisfactory anti-luck conditions. If this much can be agreed upon, why should we expect them to succeed when imposed on individualised evidence? In this chapter, my aim was to show that it is probably just a false expectation.

## 7.8 Chapter Summary

This chapter has examined a prominent modalist approach to the identify the right epistemic quality of individualised evidence. By drawing on the literature on the Gettier problem, I've shown that no modal condition on knowledge can succeed as the right anti-luck epistemic condition on individualised evidence. This also casts doubts on the prospects of modalism in applied legal epistemology. To do better, the next chapter puts forth and defends an alternative explanationist approach to individualised evidence and legal proof.

---

<sup>35</sup> Another conclusion consistent with this diagnosis is that the right anti-luck condition on individualised evidence is knowledge *simpliciter* rather than a separate modal or causal condition on knowledge. Since knowledge and veritic luck are by hypothesis incompatible, knowledge would (trivially) succeed as the right anti-luck condition on individualised evidence. I would however urge caution before drawing this conclusion. A flat-footed equation of knowledge and individualised evidence would place an even more demanding epistemological requirement on legal evidence, and it would do so on the basis an objectionable motivation (i.e., contrived versions of stylised court cases). Secondly, the putative difference between individualised and statistical evidence doesn't seem to be knowledge *per se* but rather what explains absence of knowledge (e.g., appropriate causation, sensitivity or safety). A direct appeal to knowledge hardly makes progress on the problem.

# Chapter 8

## Legal Proof as Inference to the Best Contrastive Explanation

### 8.1 Overview of the Chapter

As we have seen in the previous chapter, legal epistemology is on the rise, and special attention is now devoted to study the genuinely epistemological aspects of legal proof. This is not surprising: because it often involves notions like evidence, truth and testimony, legal proof is up in epistemology to its neck. The present chapter also contributes to legal epistemology: inspired by the explanationist analyses of knowledge previously discussed in this dissertation, I develop a novel explanationist account of legal proof. I first raise a new problem for popular modal accounts of legal proof, and show that legal proof is sensitive to the way information is presented at trial rather than to the modal profile of the information in question. This key insight motivates a different explanationist approach. By bringing out hitherto unappreciated connections between the structure of scientific explanation and the nature of legal proof, I then develop a new account of legal proof as inference to the best contrastive why-explanation. Unlike the modal account, the focus on explanation captures the importance of the way information is presented at trial. In the remainder of the chapter I articulate two motivations for the proposed account. The first is based on the most established psychological interpretation of legal proof, the story-model of legal fact-finding. The second is informed by communicative justifications of punishment: I emphasise the special role of why-explanations in favouring the repentance of the offenders and in helping the other members of the community to forgive the offenders. Overall, in this final chapter I advocate an ‘applied turn’: the proposed account of legal proof will finally yield a valuable understanding of key aspects of society such as legal evidence and criminal punishment.

### 8.2 Pritchard’s Safety Account of Legal Proof

Recall the key proof-paradoxical scenarios discussed in Chapter 6:

**Blue Bus.**<sup>1</sup> While driving back home, a vehicle suddenly hits Jane's car. In the accident, she breaks her leg. She realises that the vehicle was a bus, but she cannot discern the colour. Jane runs a quick search, and discovers that the Blue Bus Company operates 80% of buses in town. The remaining 20% are operated by the Red Bus Company. Jane sues the Blue Bus Company. Using only the evidence described here, Jane expects to win the case. After all, given the evidence, it's 0.8 probable that the Blue Bus Company is liable.

**Prisoners.**<sup>2</sup> 100 prisoners are exercising in the prison yard. Suddenly 99 of them attack the guard, putting into action a plan that the 100th prisoner knew nothing about. The 100th prisoner played no role in the assault and could have done nothing to stop it. There is no further information that can be used to settle the question of any particular prisoner's involvement. Despite the paucity of evidence, the guards pick prisoner 69 and convict him for the attack. After all, given the evidence, it's 0.99 probable that prisoner 69 is guilty.

Suppose now that the guards pick and punish a random prisoner (any will do: after all, they're all very likely to be guilty). Or suppose that liability is imposed on the Blue Bus company – also very likely to be guilty given the market share. Intuitively, these verdicts based on statistical evidence would be criticizable and epistemically improper. But add now a little twist to the case: suppose instead that an eyewitness saw a prisoner assaulting the guard, or a blue bus hitting Jane's car. In this second variation of the case, the verdict seems to meet the required standards of fairness. A satisfactory account of legal proof must account for this asymmetry in a principled and informative manner.

Pritchard's safety account of legal proof is well-equipped to explain this asymmetry. Here's a quick but faithful reconstruction of his safety account:

**Safety account of legal proof.** Total evidence E provides acceptable legal proof that  $p$  if and only if it meets a modal safety condition.<sup>3</sup>

There's a number of ways in which the safety requirement on legal proof could be met. One easy (but controversial!) way is *doxastic*: legal proof is supposed to actually generate a safe belief of guilt and liability. This is however problematic, for the relevant fact finder may not form a safe belief even if legal proof provides evidence for it. Perhaps a better and weaker way to understand the safety requirement would be to restrict it to legal proof, which is supposed to have the potential to generate safe beliefs of guilt or liability *regardless* of the mental state of the relevant fact-finder.

This modal account of legal proof has considerable explanatory power. Here is why. When based on merely statistical evidence, legal proof fails to meet a safety requirement:

---

<sup>1</sup>Redmayne (2008), Enoch et al. (2012), Smith (2018), Gardiner (2019) discuss this case.

<sup>2</sup>Originally described in Nesson (1979b). However, for the purposes of this chapter I'm adopting the presentation from Redmayne (2008).

<sup>3</sup>This account is forcefully defended by Pritchard (2018b, 2022).

much like a criminal version of the lottery paradox, in close similar possible worlds the guards could very easily (and mistakenly) punish the innocent prisoner *regardless* of the very high probability. But when based on eye-witness testimony, legal proof does meet a safety requirement and provides modal protection from error: in close similar possible worlds, the guards continue to (correctly) punish the guilty prisoner. The safety condition seems to deliver the right result. As Pritchard puts it:

In effect, the foregoing has placed a necessary condition on the evidence required to convict in criminal cases. In particular, what is required for a conviction is evidence such that, given that evidence, it cannot be an easy possibility that the defendant is wrongfully convicted. Merely having evidence that makes it highly likely that the defendant is guilty will not suffice to meet this condition, as the existence of legal lottery-style cases illustrates. (Pritchard 2018a: 117)

Because it delivers the right result, it looks like the safety account is the right account of legal proof. Indeed, at this point it's worth pausing to appreciate some key advantages of the safety account.

One key motivation for the modal account of legal proof stems from the pressure of avoiding the risk of convicting the innocent – it is vital for a well-functioning legal system to avoid such risk. If we step back and ask what standards of proof are *for*, one plausible answer to this question is that they should *minimise* the risk of convicting the innocent (in keeping with the Blackstone ratio). The safety account of legal proof understands such risk in *modal* rather than *probabilistic* terms: given the evidence adduced, legal proof should rule out an easy (modally close) possibility of wrongful conviction.

Pritchard's safety account of legal proof also complies well with descriptive projects in legal epistemology outlined in Chapter 8. Because it *describes* and *captures* the conflicted attitude of the law towards statistical evidence and the preference for individualised evidence (e.g., eye-witness testimony), it provides a powerful explanatory framework to capture existing legal practice. Moreover, it also clarifies the relevant concept of risk for legal setting: if Pritchard is right, the law in general (and legal reasoning in particular) are concerned with modal rather than probabilistic risk. Finally, Pritchard's safety account of legal proof also yields important *normative* consequences: safety is the required standard of correct for epistemically proper legal judgements.

Despite these *prima facie* advantages, there are good reasons to believe that safety is neither necessary nor sufficient for acceptable legal proof. In the next section I move two counterexamples to the necessity and sufficiency of safety for acceptable legal proof; these counterexamples pave the way for a different explanationist account.

### 8.3 Two Counterexamples, One Diagnosis

A good way to test Pritchard's account is to look at the necessity and sufficiency of safety for acceptable legal proof. Can there be unsafe and yet acceptable legal proof? Or safe and yet unacceptable legal proof? If the answer to these questions is "yes", then Pritchard's account is misguided. Indeed, there seems to be cases that suggest an affirmative answer to these questions.

Start with necessity. Let's consider two variations of the stylised court cases routinely discussed in legal epistemology:

**Hally.** While driving back home, a vehicle suddenly hits Jane's car. In the accident, she breaks her leg. She realises that the vehicle was a bus, but she cannot discern the colour. A witness, Hally, steps forward and testifies that they saw a blue bus hitting Jane's car. The tested and expected reliability of eye-witnesses is only 70%. However, unbeknownst to the court and Hally too, she suffers from colour hallucination. (Fratantonio 2021)

**Scheming Prisoner.** 100 prisoners are exercising in the prison yard. Suddenly 99 of them attack the guard, putting into action a plan that the 100th prisoner knew nothing about. Prisoner 69 is the innocent one and video evidence proves this. However, another scheming prisoner wants to make sure that prisoner 69 is punished due to long-standing resentment against him. The scheming prisoner sets out to hatch his plan to frame the assault on prisoner 69, fully ready to tamper with any piece of exculpatory evidence. However, the scheming prisoner is also lazy and, at the very last minute, he doesn't intervene.

Simplifying a little, the instances of legal proof in the two vignettes draw on i) Hally's testimony and ii) the video evidence respectively. The evidence is accurate: Hally's testimony *correctly* identifies the liable bus and the video evidence also *correctly* exculpates the prisoner. Yet, neither piece of evidence is safe: in relevantly similar close possible, Hally hallucinates and misidentifies the colour of the bus. And the scheming prisoner hatches his plan and frames prisoner 69. Does this modally close possibility of error make these two instances of legal proof unacceptable or improper? Hardly so: it would be a huge loss in accuracy if the fact finders were to disregard these two pieces of evidence. Despite being unsafe, these two pieces of evidence provide acceptable legal proof. A tentative but tempting moral to draw here is that safety is not necessary for legal proof.

On to sufficiency. Consider this other stylised court case:

**Unhinged Iggy.** Iggy Pop is on trial for public indecency after a wild gig with the Stooges, and he's accused of using and buying illegal drugs prior to the gig. Two pieces of evidence are adduced:

- (1) A witness testifies that they saw *Iggy Pop* sharing drugs with the other band members.

- (2) A drug dealer testifies that they sold drugs over the phone to *Mr. James Newell Osterberg*.

Let's grant for a moment that Iggy Pop is guilty. Crucially, the two pieces of evidence are equally accurate and modally on a par: Iggy Pop and James Newell Osterberg are one and the same person, so the close possible worlds where Iggy Pop is guilty and buys drugs are the same close possible worlds where Mr. James Newell Osterberg is guilty and buys the drugs. Accordingly, both (1) and (2) provide two equally safe instances of legal proof, but (1) has some extra significance: it's more informative and explanatory. Indeed, despite being safe, the drug dealer's testimony is not sufficient for acceptable legal proof.

As observed before, the witness's testimony seems to have special value and significance. Such extra significance doesn't seem to be captured modally: because Iggy Pop and James Newell Osterberg are necessarily equivalent, it's impossible to discriminate between them in any close (or further) possible world. Put in another way: legal proof passes the hyper-intensionality test (it introduces fine-grained distinctions between necessarily equivalent contents). As such, it can't be fully or satisfactorily captured in modal terms.

We are now in a good position to offer a *diagnosis* of the cases raised here. These cases question the necessity and sufficiency of safety for acceptable legal proof, and they do so for the following crucial reason:

**Diagnosis.** Acceptable legal proof is sensitive to *the way* information is presented at trial *regardless* of the modal profile of the information in question.

This diagnosis sheds light on what went wrong with safety in the cases raised above. Despite the close error possibility, the pieces of evidence adduced in **Hally** and **Scheming Prisoner** explain why the defendant are guilty or innocent. Differently, despite being safe, the drug dealer's testimony failed to explain why Iggy Pop was guilty.

The moral to draw from these cases is that the right account of legal proof is explanationist rather than modal.<sup>4</sup> In the next section, I move on to develop such an account.

## 8.4 Legal Proof as Inference to the Best Contrastive-why Explanation

While initially plausible, the modal account has been found wanting: because of the hyperintensional structure of legal proof (and proof more generally), modal conditions are simply ill-suited to capture the nature of legal proof. This observation calls for a different

---

<sup>4</sup>This diagnosis differs from Frantonio's general pessimism about the feasibility of *any* epistemic condition on legal proof (Frantonio 2021). The problem is not with epistemic qualities in general. Rather, it is restricted to modal conditions: modal conditions specifically fail to elucidate legal proof. I am making a weaker point: legal proof can't be fully captured in modal terms, but another explanationist condition may still succeed.

approach: what the foregoing shows is that legal proof is a matter of explanatory relations. The aim of this section is to sketch out such an account.

Explanationist approaches to legal proof have been a minority in the legal epistemology literature.<sup>5</sup> At a minimum, these approaches must specify the *type of explanation* in question and its *key features*. Let's take each in turn.

According to the explanationist account of legal proof defended here, the relevant type of inference is *abductive* (also called “inference to the best explanation”; see Harman (1965 and Lipton 1991 for a seminal work on the topic). Roughly, an inference to the best explanation consists in selecting the hypothesis that best (or better) explains the data. Suppose for example that your refrigerator suddenly stopped working along with other appliances in your apartment, and the lights in the houses nearby are also off (this being the *explanandum*). Different candidate explanations (*explanantes*) are offered: perhaps someone pulled a prank on the neighbours (*explanans 1*), perhaps the appliances stopped working because they all have the same defect (*explanans 2*), or perhaps there's a power outage in your part of town (*explanans 3*). Clearly, the hypothesis that best explains the data is the third one: the a general power outage better explains why the appliances stopped working and also why the lights are off in the nearby houses. This is an instance of abduction: the comparatively better hypothesis is selected on the basis of explanatory virtues. Judicial proof also follows this procedure: given the evidence adduced to trial, a positive or negative legal verdict should align with the better explanation of the evidence adduced.

Focussing on abductive inference also helps to shed light on the key features of the explanations that are relevant for judicial proof. Plausibly, these are *why-explanations*: indeed, the evidence adduced to trial should explain *why* the defendant is guilty in a criminal trial. More precisely:

Total evidence E meets the beyond reasonable doubt standard (if and) only if E provides the best satisfactory, comprehensive and robust *de re* why-explanation of the defendant guilt.

This is a new and informative necessary condition on legal proof (I remain neutral on any sufficiency claim). To be maximally clear, I now clarify each key term underlined in the formulation with a special emphasis on the key features of the.

- *Total evidence E* refers to all the admissible evidence adduced to the case (e.g., eye-witness testimony, fingerprints, video evidence).
- An explanation is *satisfactory* if it's acceptable by a normal rational agent, *comprehensive* if it covers most of or all the evidence adduced to case and *robust* if it displays internal coherence.

---

<sup>5</sup>Allen (2008; 2013) and Allen and Leiter (2001) defend such approaches. However, their motivations are different: because they draw on a naturalised *a posteriori* legal meta-epistemology, Allen (and Leiter) set out to understand legal proof without any appeal to *a priori* methodologies based on hypothetical cases. My motivations are different: I defend explanationism in legal epistemology because of the problems of modal accounts with ensuring the right explanatory connections between the evidence adduced and the verdict.

- Why-explanations address relevant why-questions raised during various stages of legal fact-finding. Why-explanations come in different degrees of specificity: they target why-questions that are both narrow (e.g., why was the defendant present on the crime scene at the time of the crime?) and broad (e.g., why should the defendant be found guilty?). Crucially, acceptable why-explanations are *de re* – they refer directly to the defendant rather than to a wider reference class.

For sake of vividness, consider how the proposed account deals with this toy example. Suppose that Jones is accused of murdering Smith. The evidence presented is the following: Jones owed money to Smith, someone saw the two of them arguing in a bar and Jones' hunting knife was found on the crime scene. Each piece of evidence provides a satisfactory, comprehensive and robust why-answer to key why-questions. First, motive: why would Jones murder Smith? Because he owed him money and they had an argument. Second, the murder weapon: why was the knife present on the crime scene? Because it was used by Jones to kill Smith. Finally, the verdict: why should Jones be found guilty? Because he owed money to Smith, he argued with him in a bar and his knife was found on the crime scene.

We now have a more detailed picture of the relevant explanations for legal proof – an inference to the best *de re* why-explanation. There's one final element to add, and it pertains to the *contrastive* nature of explanation. The point is familiar in the philosophy of science literature on the nature of scientific explanation. Indeed, why-explanations never explain just one fact; rather, they explain one fact *rather than*, or *in contrast* with, another.<sup>6</sup> As Bas Van Fraassen (1980: 127) famously puts it: “The correct general, underlying structure of a why-question is... Why (is it the case that) P in contrast to (other members of) X where X, the contrast class, is a set of alternatives”. For example, to ask why did the water boil when it reached 100 degrees is to implicitly ask why did the water boil when it reached 100 degrees *rather than* a different temperature, or why did the water boil *rather than* turn into ice.

Crucially, the contrastive nature of why-explanations is found also in the structure of legal proof. Plausibly, legal proof never proves a proposition *p* but rather proves a proposition *p* rather than a different proposition *q*. A piece of compelling evidence against the defendant doesn't just prove why they're guilty: instead, it proves why they are guilty rather than innocent.

This last point bears special emphasis because it clarifies a general and important feature of standards of proof. Legal proof is a matter of establishing facts at a level of *comparative plausibility*: the claims of the prosecution and the defence are competing hypotheses that purport to explain the facts at issue, and the relevant fact-finders (a judge or a jury) reason to select the hypothesis that *better* explain the evidence adduced. The probabilistic and modal accounts of legal proof struggle to account for this feature of the standards. The probabilistic account assumes a very demanding probabilistic reasoning that is not available to the average fact-finder. And similarly, the modal account can't accommodate the

---

<sup>6</sup>See Shan (2019) for extensive discussion on the contrastive nature of explanation.

graded and comparative assessment of the hypotheses offered in court (modal reasoning is also equally unavailable to the average fact-finder). The explanationist account sketched out here does a better job at capturing this feature of standards of proof, and, as such, it should be preferred.

Let's take stock. I have defended an explanationist account of legal proof, which amounts to an inference to the best contrastive *de re* why-explanation. After outlining the account, I'd like to provide some motivations for it. In the next section, I move on to offer three arguments in its favour.

## 8.5 The Linguistic Argument

One initial way to bolster the case for an explanationist account of legal proof focusses on *linguistic patterns* surrounding judicial proof and impositions of criminal punishment and liability. In a nutshell, the linguistic argument proceeds as follows:

- $P_1$  If linguistic evidence supports the account of legal proof under consideration, then we have good reasons to take the account under consideration to be correct.
- $P_2$  Linguistic evidence supports the explanationist account of legal proof.
- C: We have good reasons to take the explanationist account of legal proof to be correct.

Before setting out this crucial pieces of linguistic evidence in support of the proposed explanationist account, I will make the assumptions underlying the linguistic argument a bit more explicit. Let's zoom in on the first premise of the linguistic argument. The key idea here is that inquiry into linguistic patterns surrounding legal proof advances our understanding of legal proof itself. The first premise of the argument is intuitively plausible and also accepted elsewhere: inquiry into the nature of other (epistemic) states also proceeds by focussing on linguistic patterns and ascriptions.<sup>7</sup>

Having clarified the methodological assumptions underlying premise 1 of the linguistic argument, we can now move on to the linguistic evidence that favours the account. Consider the following:

**Utterance 1.** “The evidence adduced to trial proved that the defendant was innocent but didn't explain why they were innocent.”

---

<sup>7</sup>The focus on knowledge ascriptions in epistemology is long-standing (see for instance Stanley 2005 and Gerken 2017). A similar methodological approach can be found also in the literature on understanding. Egler (2021: 6062) nicely makes the point as follows: “The central assumption underlying this approach is that our evaluations of the sort ‘S understands why *p*’ can shed light on which conditions S must satisfy in order to have interrogative understanding of *p*. Thus, it is generally assumed that we can gain explanatory purchase on what is required to understand, say, why crows fly by investigating the circumstances under which we ascribe understanding of why crows fly to others or to ourselves.” In the same vein, the linguistic argument purports to advance the understanding of legal proof by focussing on the linguistic patterns surrounding legal proof.

**Utterance 2.** “The jury convicted the defendant, but they didn’t understand why they were guilty.”

**Utterance 3.** “Total evidence E proves *p*, but it doesn’t explain *why p*.”

Crucially, these three utterances all sound infelicitous and therefore in tension with acceptable legal proof. Take for instance **Utterance 1**: in this case, it’s unclear to what extent the evidence adduced can meet the standard of proof under consideration without providing any explanation of the defendant’s guilt. On to **Utterance 2**: in this case too, we’d tent to criticise the jury’s decision because of the lack of explanatory connections between the evidence and the verdict. Finally, **Utterance 3** also points at a tight link between successful explanation and legal proof: successful explanation looks like a necessary condition on legal proof.

More utterances in support of the explanationist account of legal proof can be generated. The 3 I have provided above can serve as a general recipe: the broad idea is to pair successful legal proof with absence of explanation. Doing so shows how there can’t be the former without the latter. The converse also holds: satisfactory explanation seems to track successful legal proof. Consider:

**Utterance 1.** “The evidence adduced to trial proved that the defendant was innocent and explained why they were innocent.”

**Utterance 2.** “The jury convicted the defendant and understood why they were guilty.”

**Utterance 3.** “Total evidence E proves *p* and explains *why p*.”

In the revised versions of the utterances, acceptable legal proof aligns with satisfactory explanations. This lends further support to the explanationist account of legal proof defended here. Even if not conclusive, linguistic evidence provides us with good reason to think that the explanationist account of legal proof is on the right track.

## 8.6 The Psychological Argument

The second argument I want to offer in favour of the explanationist account of legal proof moves from some key observations regarding the psychological processes and reasoning underlying legal fact-finding. As with the linguistic argument, for sake of clarity it’s worth adopting the standard form:

$P_1$  If a psychological model of legal fact-finding supports the account of legal proof under consideration, then we have good reasons to take the account under consideration to be correct.

$P_2$  The “story model” model of legal fact-finding supports the account of legal proof under consideration the explanationist account of legal proof.

C: We have good reasons to take the explanationist account of legal proof to be correct.

This time, most of the action is on the second premise. Here, more clarification is needed: what is the story model of legal fact-finding, and how does it support the explanationist account? In the remainder of the section, I will give an answer to both questions.

Pennington and Hastie's psychological model of legal reasoning is sometimes called the "story" or "scenario" model.<sup>8</sup> Their basic methodological approach is the following: to advance our understanding of legal proof it's worth looking at how agents actually reason when they consider the (big) amount of evidence adduced in court. As such, the story model is a descriptive theory (a psychological account) of legal fact-finding.

Central to their model is *active story reconstruction*: fact-finders strive to impose a narrative structure to the evidence presented at trial. In essence, they try to impose a "narrative story organisation" to the information available in court (Mackor and Van Koppen 2021: 216).

But what are stories? Simply put, stories are narratives that can be broken down into elements and episodes. In turn, episodes consist of specific elements like initiating event, a psychological response, a goal or an aim, an action and their consequence, and so on. For instance, the following episode can be thought of a rather simple story: a husband has a fight with his wife (*initiating event*), which makes him very upset (*psychological response*). Because he wants to hurt her (*goal*), he hits her (*action*), which then causes her death (*consequence*).<sup>9</sup> The example shows that episodes have a common structure: elements are chronologically ordered and partly connected on the basis physical and mental causal relationships. Stories are, overall, a hierarchy of episodes. This is what the story model of legal fact-finding predicts: as an established psychological theory, it deserves to be taken seriously.

On to the second question: how does this model support the explanationist account of legal proof defended here? Crucially, the exercise of active story reconstruction and the imposition of a narrative structure proceeds via instances of inference to the best explanation. In Pennington and Hastie's terminology, stories are nothing over and above the hypotheses that best explain the evidence adduced at court. As a result, the story model supports the explanationist account of legal proof defended here: indeed, the explanationist account provides robust philosophical foundations for the story model.

The tight and important connection between the story model of legal fact-finding and the explanationist account of legal proof has been noted also elsewhere. Here's a telling quote from Mackor and Van Koppen:

The scenario theory holds that in criminal cases explanations should be structured as scenarios consisting of episodes with the elements described by the story model, such as an initiating event, psychological response, goal, action, and consequences. In that manner IBE and the story model are combined.

---

<sup>8</sup>Pennington and Hastie (1991, 1993).

<sup>9</sup>Example borrowed and reworded from Mackor and Van Koppen (2021: 216).

First, the elements take the form of hypotheses that together make up a scenario, analogously to how scientific hypotheses together make up a scientific theory. Since in a good scenario the elements are chronologically ordered and partially causally connected, they are not a random bunch of hypotheses, but form a more or less coherent whole, a holistic story about what may have happened. (Mackor and Van Koppen 2021: 122)

Overall, a very accepted and established psychological theory like the story model of trial supports the explanationist account of legal proof defended here. But even if admittedly powerful, psychological evidence is never conclusive: accordingly, the limits of the psychological argument should be pointed out. I do so to conclude the section.

The psychological argument is a *descriptive* argument: the best it can do is to show that the explanationist account is psychologically feasible – it captures the way fact-finders reason when they consider the evidence adduced to court. However, it remains silent on the normative aspects of legal proof. It's easy to think of cases in which the narrative reconstruction is based on explanatory powerful but racist or controversial stereotypes.<sup>10</sup> In this cases, the explanatory demands are met, and yet the conviction seems to be improper (from a moral or perhaps even epistemological point of view). Accordingly, different theoretical commitments and further interests into the normativity of legal proof can provide enough reason to be suspicious of the psychological argument.

## 8.7 Coda: Norms of Criminal Conviction and Explanation

The explanationist account of legal proof defended here has noteworthy consequences for another important case study in legal epistemology – the epistemic normativity of criminal convictions. For example, a modal account of legal proof yields the following epistemic norm of criminal conviction:

**Safety Norm of Criminal Punishment (SN).** Convict the defendant if and only if, given the admissible evidence, you *safely* believe that they are guilty. (Pritchard 2018b)

Motivations for the safety norm mirror the motivations for the safety account of legal proof. Recall the prisoners example discussed in the previous chapters (Nesson 1979a). Suppose that the guards pick and punish a random prisoner (any will do: after all, they're all very likely to be guilty). Intuitively, this criminal conviction based on statistical evidence would be criticizable, and epistemically improper. But add now a little twist to the case: suppose instead that an eyewitness saw a prisoner assaulting the guard. In this second variation of the case, the conviction seems to meet the required epistemic standard for criminal punishment. Pritchard's safety norm is well-equipped to explain why.

---

<sup>10</sup>See Gonzalez Rose (2021) for a representative sample of cases.

When based on merely statistical evidence, the belief of guilt is unsafe: much like a criminal version of the lottery paradox, in close similar possible worlds the guards could very easily (and mistakenly) punish the innocent prisoner. But when based on eye-witness testimony, the belief of guilt is safe: in close similar possible worlds, the guards continue to (correctly) punish the guilty prisoner. Because (SN) delivers the right result, it looks like safety can successfully elucidate the epistemic normativity of criminal punishment.

However, we have seen how safety is ill-suited to elucidate the structure of legal proof. The key point here is that modal conditions fail to ensure the right explanatory connections between the evidence and the verdict. Moreover, as we it's been showed above, safety is neither necessary nor sufficient for legal proof.

These considerations motivate a different norm of criminal punishment, one that pays heed to the importance of explanatory connection. Such norm is an understanding norm:

**Understanding Norm of Criminal Punishment (UN).** Convict the defendant if and only if, given the admissible evidence, you understand *why* they're guilty.

This understanding norm drops out naturally from the explanationist account of legal proof defended here. Notice also that the understanding norm nicely diagnoses what's wrong with merely statistical evidence: in fact, merely statistical evidence fails to explain why the defendant is guilty. For these very reasons, the right norm for criminal conviction must make reference to understanding instead of other epistemic states such as safe, knowledgeable or justified belief.

## 8.8 Chapter Summary

In this chapter, I have expanded on the problems of modal accounts of legal proof by pointing out the hyperintensional features of legal proof. Once these features are brought into focus, the modal accounts are essentially limited. To do better, I sketched out a different explanationist account. According to this account, legal proof is an inference to the best contrastive de re why-explanation. I have developed two arguments in support of this account (a linguistic argument and a psychological argument). Finally, I derived an understanding norm of criminal convictions from the explanationist account defended here.

# Chapter 9

## Conclusions

### 9.1 Knowledge and Safety

**Recap.** The starting point of this dissertation was found in the contrast between two competing approaches to the analysis of knowledge: modalism and explanationism. The former employs a familiar *intensional* possible world framework made popular in other areas of philosophy, and it equates knowledge with a belief which is true in a relevant set of possible worlds. By contrast, the latter focusses on the *hyperintensional* notion of explanation. According to explanationism, knowledge amounts to a belief hold on the basis of the right explanation (for example, hold because it's true).

I have assessed the two most developed versions of modalism and explanationism: safety-theoretic accounts of knowledge and kairetic explanationism. This dissertation has advanced the debate in the following ways:

- *In defence of Modalism.* Thanks to a better construal of the safety condition, counterexamples to the necessity and sufficiency of safety for knowledge can be defused. Both environment-relative safety and modal proper functionalist readings of the safety condition successfully achieve that.
- *Against Kairetic Explanationism.* The most recent and best developed version of explanationism – kairetic explanationism – fails to deliver as satisfactory anti-Gettier condition. Indeed, this version of explanationism is committed to bad closure failures (in *de dicto* and *de re* barn-style case), and it also falls prey to Gettier-style counterexamples.

**Directions for further research.** As explained in Chapter 5, a hitherto unexplored and promising way of developing safety consists in deriving a safety condition from a proper functionalist condition. An even better way to derive safety would be to construe proper function as a specific type of aptness: roughly, knowledge would be equated with a belief which is true (accurate) because of/in virtue of an etiological proper function condition. In future work, I plan to clarify this fit between truth and proper function in more precise terms. A natural place to draw inspiration from is virtue reliabilism: the fit between

truth and proper function can be understood in *explanatory* terms like Greco's theory of knowledge or in *dispositional* terms like Sosa's virtue epistemology.

## 9.2 Knowledge and Legal Proof

**Recap.** The contrast between modalism and explanationism can be found also in legal epistemology – the branch of applied epistemology concerned with legal proof. When it comes to applied legal epistemology, in this dissertation I have objected to modalism and I have defended explanationism. First, I pointed out the inadequacies of modal accounts of legal proof by divorcing anti-luck conditions on knowledge from anti-luck conditions on individualised evidence. Then, I provided a diagnosis of the inadequacies of modal accounts of legal proof by showing that legal proof is hyperintensional. Finally, I sketched out an alternative explanationist account of legal proof where inference to the best explanation takes centre stage. I argued in favour of this account by drawing on linguistic and psychological considerations.

**Directions for further research.** In future work, I plan to better bring out the analogies between scientific explanation and legal proof: I set out to make the contrastivist features of legal proof more precise; doing so will strengthen the connection between abduction and legal proof. I also intend to investigate what the account predicts when tested against the civil and criminal versions of the proof paradox. Finally, I plan to extend the explanationist account to cases of expert testimony in courts, and show how my proposed account can capture such crucial legal phenomenon.

I don't pretend to have provided exhaustive answers to the questions raised in this dissertation, and I hope to make further progress on all these questions in my future work.

# Appendix A

## Swampman Cases

In Chapter 5, I developed a proper functionalist version of safety:

**Safety as Modal Proper Functionalism.** S's belief that  $p$  is safe if and only if in all the possible worlds in which S continues to form her belief in the target propositions  $p$  via the same properly functioning belief-forming method  $M$  employed in the actual world (sub-condition M) and in the same appropriate cognitive environment  $E$  occupied in the actual world (sub-condition E), the belief continues to be true.

To bring out the key elements of this novel version of safety, I also compared it and contrasted with other (and more standard) versions of the safety condition. But what about standard proper functionalism? Since it contributes to provide a satisfactory criterion of individuation of belief-forming methods, appropriate environments and modal robustness, the concept of proper function is doing a lot of the explanatory work here. It is therefore only natural to wonder whether – and to what extent – this new formulation of safety is also substantially different from standard versions of proper functionalism.

Because it's one of the key objections faced by standard versions of proper functionalism, the case of Swampman represents the most meaningful test bench for the proper functionalist safety condition. A good way to bring out the main differences between these two accounts is to see how each deals with the much debated Swampman case, which goes as follows:

**Swampman.** Suppose lightning strikes a dead tree in a swamp, and someone is standing nearby. Their body is reduced to its elements, while entirely by coincidence (and out of different molecules) the tree is turned into their physical replica. Such replica, The Swampman, moves exactly as the original did; according to its nature it departs the swamp and begins to form true beliefs. (Sosa 1993: 53)

The vignette spells trouble for standard versions of proper functionalism, most notably the proper functionalist account of knowledge defended by Plantinga (1993). In fact, as intuitions would have it, Swampman's true beliefs amount to knowledge; however, since

it's the result of a cosmic accident, Swampman lacks any design – conscious, natural or otherwise. Given the knowledge intuition in the absence of proper function, the case suggests that proper function is not necessary for knowledge.

In response, proper functionalists have denied that the Swampman scenario is genuinely possible and also that Swampman's beliefs display any positive epistemic status (knowledge, justification, entitlement or what have you).<sup>1</sup> But this line of response is not problem free: the knowledge intuition persists, and advocates of proper functionalism do sometimes concede that the Swampman case is coherent, conceivable, and therefore possible.<sup>2</sup>

These considerations pave the way for the first stark difference between standard versions of proper functionalism and the proper functionalist safety condition offered in Chapter 5. Crucially, the advocates of the proper functionalist safety condition can wholeheartedly concede that Swampman is genuinely metaphysically possible. In addition, they also have the theoretical option to deny or defend the knowledge intuition depending on their preferred reading of the modal profile of Swampman's beliefs. Given the central safety-theoretic focus on relevantly similar close possible worlds, it is worth breaking down the discussion of Swampman into two distinct parts: we can assess the case either indirectly from the perspective of the actual world or directly by taking Swampman-style worlds as actual. Let me take each in turn.

From the perspective of the actual world, Swampman is a genuinely conceivable but nonetheless far off possibility. After all, Swampman-style worlds are too dissimilar: they require a substantial change in the biological history, laws of nature, belief-forming methods and environments of the actual world. More to the point, from the perspective of the actual world Swampman is on a par with the radically deceived victims of skeptical scenarios such as a brain-in-a-vat which has always been envatted. Even if genuinely possible, both the Swampman and the brain-in-a-vat scenarios require a considerable change from the actual world.<sup>3</sup> But since all that matters for the safety condition on offer here (and for the safety condition in general) is the swath of relevantly similar close possible worlds, these substantially different – and therefore far off – possible worlds do not matter. If we focus on the actual world, the Swampman case doesn't constitute a genuine counterexample to the proper functionalist version of safety: the case is possible but too dissimilar and, as such, ultimately irrelevant.

---

<sup>1</sup>Commenting on the case, Plantinga (1991: 206) writes: "I haven't any reason to think this possible; I'm not positive that it is impossible; but I am inclined to think it is, and hence don't consider it a counterexample to my account". Similarly, Graham also denies that Swampman is genuinely possible: in fact, he explicitly sets out to "deny the metaphysical assumption that Swampman has a mind" (Graham 2012: 467).

<sup>2</sup>As Sosa notices (1993: 56), Plantinga seems to concede that there's something intuitively coherent about the Swampman thought experiment (Plantinga 1991: 206). Graham (2014b: 30) appeals to biological cases of macromutations to offer a more realistic Swampman case. See Simion (2016b: 2.2) for helpful discussion.

<sup>3</sup>For example, Sosa (1999: 147) emphasises the anti-sceptical import of the safety condition by pointing out the substantial dissimilarities between the actual world and the far-off sceptical worlds. Given the clear nomological contrast between the actual world and Swampman-style worlds, the dissimilarities between these two possible worlds should be even easier to detect regardless of the alleged anti-sceptical import of the safety condition.

But suppose that instead we take the Swampman's world as actual: wouldn't this undermine the proper functionalist construal of safety developed in this chapter? After all, Swampman would lack any intentional or natural design plan, and, accordingly, Swampman's beliefs would fail to meet the proper functionalist criterion of individuation of belief-forming methods previously outlined. Just like standard proper functionalist accounts of knowledge, the proper functionalist version of safety would also fail to accommodate the intuition that Swampman knows, and this is clearly bad news for both standard proper functionalism and the proper functionalist version of safety on offer here.

Matters are delicate, but let's begin to note that Swampman's belief-forming methods are *prima facie* functioning properly: even if he lacks any robust biological history, his perceptual capacities seem to be meeting some proper function condition – there's no clear malfunction. The same goes for his cognitive environment: given the absence of misleading features, Swampman's environment is appropriate for the belief-forming methods under consideration. Once these two parameters are kept fixed, Swampman does continue to form true beliefs across relevantly similar Swampman-style close possible worlds. These considerations help to capture the intuition that Swampman knows. After all, the belief-forming method and the environment under consideration both meet a proper functionalist condition: as such, Swampman's true beliefs seem to be safe and, accordingly, they are in the market for knowledge.

Yet, Swampman lacks any history or design, and so do his belief-forming methods. However, depending on where the source of the design is located, more can be said about the (lack of) design of Swampman's belief-forming methods. Under a strictly naturalistic conception of design, functions are selected because of a rather long biological history of success (Millikan 1984; Graham 2012). However, functions need not be individuated by reference to a long history of biological benefit. Suppose for instance that Simion (2016b: Section 4.2) is right, and what ultimately matters for function acquisition is some positive epistemic rather than biological (and ultimately historical) benefit. Accordingly, as long as Swampman forms true beliefs on the basis of an epistemically beneficial method like perception, his belief-forming methods acquire a function and are therefore not only functioning properly, but they are functioning properly according to a selected function that aims at truth. Once the historical requirements on the source of the design plan are relaxed, Swampman's beliefs can meet the proper functionalist safety condition.<sup>4</sup>

Taken together, these considerations bring out important differences between standard versions of proper functionalism and the proper functionalist version of safety defended here. Unlike standard versions of proper functionalism, the proper functionalist version of safety grants that Swampman is genuinely possible. This is the first important difference. Moreover, depending on the preferred reading of the modal profile of Swamp-

---

<sup>4</sup>Sosa (1993: 58) argues that such historical requirement could be abandoned entirely: building on his influential conception of intellectual virtue, he proposes a weaker notion of proper function that makes no reference to an evolutionary or divine design. While this notion of proper function is different and more similar to a virtue-theoretic ability condition, it remains consistent with the strategy to elucidate safety that I pursue in this thesis.

man's beliefs, advocates of the proper functionalist version of safety can reject the knowledge verdict by dismissing Swampman-style worlds as too dissimilar or accommodate the knowledge intuition by relaxing the historical requirement on the source of function of belief-forming methods. This is the second important difference. Either way, the proper functionalist version of safety diverges from more standard versions of proper functionalism. Given the emphasis on the modal profile of the beliefs under consideration (usually unavailable to standard versions of proper functionalism), this should not be surprising. Overall, this new proper functionalist version of safety displays important differences also with the standard versions of proper functionalism developed in the literature.

# Appendix B

## Epistemic Supervenience and the Analysis of Knowledge

### Overview

The central objections to knowledge-first epistemology, since its arrival on the scene more than 20 years ago, are now mostly well known. While many of these objections have been methodological, other criticisms have been more substantive. It is fair to say that, at present, all of the main substantive objections to knowledge-first epistemology have received sophisticated responses on behalf of knowledge firsters – all with the exception of a very new kind of objection called the *supervenience objection* (e.g., Pritchard and Kallestrup 2017). The supervenience objection challenges knowledge firsters' metaphysical credentials by maintaining that the view is incompatible with uncontroversial supervenience requirements on epistemic properties. This appendix charitably constructs this hitherto entirely unaddressed objection on behalf of knowledge-first critics and shows that, ultimately, the objection does not stick. The upshot is a clear appreciation of both why supervenience does not make any trouble for the knowledge-first project, and moreover, why knowledge firsters have stronger metaphysical credentials than even its proponents have thus far appreciated.

### Introductory Remarks

The knowledge-first revolution in epistemology takes knowledge as the most basic notion of epistemological theorizing. Since its arrival on the scene twenty years ago, the knowledge-first approach has marked a paradigm shift in the discipline: recent results in epistemology and neighbouring areas such as the philosophies of language and mind compellingly show that the knowledge-first approach is a thriving research programme. However, its opponents have not been idle: many diverse and sophisticated objections have been levelled against the key claims advanced in the knowledge-first camp (most notably, the equation of evidence and knowledge, the knowledge norm of assertion and

practical reasoning, knowledge as the most general factive mental state). Thus far, most of these criticisms have received equally sophisticated responses on behalf of knowledge firsters – all with the exception of an entirely novel kind of objection, the supervenience objection. This objection has to do with the metaphysics of knowledge, and aims to show that putting knowledge first essentially clashes with elsewhere accepted supervenience constraints placed on epistemic properties. The primary aim of this appendix is to discuss this hitherto unaddressed objection and argue that it fails: knowledge firsters have in fact the resources to meet the supervenience challenge.

Here is the game plan. I begin by introducing the notion of supervenience and by explaining its relevance to epistemology. Next, I charitably reconstruct the supervenience objection. For sake of clarity, I split the objection into three different parts: the base properties problem, the epistemic superdupervenience demand and the failure of priority. Then, I proceed to show that each part ultimately fails. To anticipate: the base properties problem is addressed by insisting that knowledge is a *sui generis* factive mental state. The epistemic superdupervenience demand is fairly more complex, and therefore deserves more extensive treatment. Hence, I develop two different lines of response – a refusal and an acceptance line. The former rejects the requirement in light of implicit reductionist assumptions; the latter provides an epistemic superdupervenience in terms of *identity* and *metaphysical grounding*. Finally, I defuse the last strand of the objection by distinguishing between conceptual and metaphysical priorities of knowledge. My conclusion will be that knowledge firsters have strong metaphysical credentials, and that supervenience does not cause any trouble to their project.

## Epistemic Supervenience

Epistemic properties do not float free. Rather, they depend on more basic non-epistemic properties in virtue of which they are instantiated. For instance, reliabilists notoriously hold that a belief is justified when it has been produced and sustained by a reliable process (Goldman 1979). What this renown example shows is that, intuitively enough, the epistemic property *being justified* depends on a different, more fundamental property, (i.e., *being produced by a reliable process*).<sup>1</sup> This somewhat loose dependence relation is better captured by supervenience: a set of properties A supervenes upon another set B in case no two things can differ with respect to A-properties without also differing with respect to their B-properties. In a slogan form, no difference in A-properties without a difference in B-properties. Before moving on to the epistemic case and see how supervenience is supposed to cause problems for knowledge firsters, it is worthwhile to dwell briefly on such relation and its relata. Hopefully, doing this will clarify the forthcoming discussion.

As a metaphysical relation, supervenience has the following formal features: it is transitive, non-symmetric (i.e., compatible with both symmetry and asymmetry) and reflexive. However, for the purpose of the present appendix, it will be assumed to be asymmetric.

---

<sup>1</sup>I follow the standard practice and refer to *properties* in italics and to concepts in small caps.

In addition, it also has a specific modal force. Despite a variety of options,<sup>2</sup> it will be assumed to be Strong Individual: that is, quantifying over individuals across different possible worlds. Borrowing from Kim (1993), here what a formulation of Strong Individual Supervenience looks like:

A-properties *strongly supervene* on B-properties if and only if for any possible worlds  $w_1$  and  $w_2$  and any individuals  $x$  in  $w_1$  and  $y$  in  $w_2$ , if  $x$  in  $w_1$  is B-indiscernible from  $y$  in  $w_2$ , then  $x$  in  $w_1$  is A-indiscernible from  $y$  in  $w_2$ .

So much for supervenience relations. Next, onto supervenience relata. Here, matters are not very straightforward, given the difficulties of identifying the epistemic and non-epistemic domains with sufficient accuracy. However, intuitive examples will do at this preliminary stage. Let supervenient epistemic properties be those properties epistemologists are typically concerned with: *knowledge, justification, defeasibility, evidence, warrant, rationality*. These properties certainly vary with respect to their degree of complexity: while *justification* comes in at least two varieties, doxastic and propositional, *warrant* entails non-defeated belief, which requires a number of (complex) conditions to be met. However, it should be conceded that they all represent genuine instances of epistemic properties. Subvenient non-epistemic properties, on the other hand, are not as easily identified. Still, given that epistemology is widely taken to be a normative enterprise, non-epistemic properties will be henceforth identified with descriptive/non-evaluative properties such as mental properties, natural properties, psychological states and processes. This will facilitate the discussion, and narrow down our focus to a reasonable extent. Accordingly, we can formulate epistemic supervenience as follows:

**Epistemic Supervenience (ES):** Necessarily, whenever an agent A has a normative epistemic property E, then A also has a collection of descriptive non-epistemic properties N such that, necessarily, any other agent A\* with N also has E.<sup>3</sup>

So formulated, **(ES)** states an asymmetric covariance relation holding between epistemic properties and their non-epistemic base. Nevertheless, when used as dialectical weapon, epistemic supervenience raises a fundamentally *two-fold* challenge. Indeed, a suitable non-epistemic base must be identified. However, identifying the correct subvenient properties will not suffice in itself: in fact, what is additionally required is a ‘*superdupervenience*’, i.e. a fine-grained metaphysical relation that explains why the covariance pattern of supervenience holds in the first place, and it is not simply a further inexplicable brute fact.<sup>4</sup>

Before explaining how exactly **(ES)** causes trouble to knowledge firsters, it is worth spending a few words to characterise their lot. Knowledge-first epistemology has recently

<sup>2</sup>See McLaughlin and Bennett (2021: §3.1) for an overview.

<sup>3</sup>Readapted from Pritchard and Kallestrup (2017). See also Turri (2010) for a comprehensive overview of the role of supervenience in epistemology. Van Cleve (1985) and Lyons (2009: chapter 2) both avail themselves of supervenience arguments in epistemology.

<sup>4</sup>The term was firstly introduced by William Lycan, but see Horgan (1993: 566) for a formulation of the superdupervenience requirement. More on this in §4.

gained momentum among the practitioners of the discipline: in light of the failed attempts to analyse knowledge into more basic components started with Gettier's famous short piece (Gettier 1963), knowledge firsters set out to take it as a primitive and then use it to explain other epistemic, linguistic and psychological phenomena (such as evidence, justification, assertion and action). Williamson's *Knowledge and Its Limits* represents the first seminal attempt to reverse the standard order of explanation implicitly assumed by analytic epistemologists since Gettier's appendix; however, in spite of Williamson's ambitious proposal, knowledge firsters need not (and typically do not) embrace each of Williamson's theses. Putting knowledge first mirrors a broader methodological approach and the membership to a novel research programme rather than a complete acceptance of Williamson's own view(s).<sup>5</sup> Be that as it may, a very significant advantage of knowledge-first epistemology is to avoid the long series of counterexamples that plagued the previous analytical project, and to pave the way for novel conceptions of traditional epistemic notions such as evidence, and justified belief.<sup>6</sup> It thus holds far-reaching potential, and, accordingly, such potential motivates a persuasive defence from the supervenience objection.

Importantly, this objection is particularly significant, and thus worthy of special attention. Firstly, while many of the criticisms levelled against knowledge-first philosophy have already been explored, this one has been thus far overlooked. Secondly, the supervenience objection aims to cast doubts on knowledge-first metaphysical credentials – i.e., it purports to show that putting knowledge first is *metaphysically inconsistent*. If it hits the mark, the objection threatens the viability of the knowledge-first project, since it would show that putting knowledge first clashes with uncontroversial supervenience requirements on epistemic properties. Finally, it is worth noticing the potentially wide target of the objection. First off, the objection applies to Williamson's knowledge-first views, as we will see in the next sections. However, far from being a mere quibble at Williamson's own knowledge-first philosophy, the objection generalises against *each* knowledge-first project that explains epistemic (or evaluative) properties in terms of knowledge. These projects include recent accounts of justification in terms of knowledge or would-be knowledge (Sutton 2007, Bird 2007, Reynolds 2013, Ichikawa 2014, Kelp 2017, Miracchi 2015, Simion 2019, Littlejohn forthcoming) or knowledge-based accounts of reasons and rationality (Littlejohn 2017).

For ease of exposition, it will be useful to split the objection into three parts: (1) *the base properties problem*, (2) *the epistemic superdupervenience demand* (3) *the failure of priority*. In the remainder of the appendix, I set out to outline every strand of the objection and show that each ultimately fails. As already anticipated, my strategy hinges on two relatively straightforward tenets of the knowledge-first approach, namely that knowledge is a *sui generis* factive mental state (Williamson 2000: chapter I) and that knowledge exhibits different kinds of priority (Ichikawa and Jenkins 2017). Accordingly, the plan is as

---

<sup>5</sup>Carter, Gordon and Jarvis (2017: 7) also stress this point.

<sup>6</sup>For detailed overview of the history of these counterexamples, see Shope (1983). For an alternative and interesting line of narrative, see Antognazza (2015: §2) and Dutant (2015). For a critical take on knowledge-first philosophy, see McGlynn (2014).

follows. Given that the base properties properties and the epistemic superdupervenience demand are tightly connected, I will address them consecutively in §3 and §4. Conversely, I will devote a final stand-alone section to the failure of priority, since I consider its treatment as fairly independent from the other two strands of the objection.

## The Base Properties Problem

The first strand of the supervenience objection has to do with the non-epistemic base properties upon which relevant epistemic properties are required to supervene. In fact, the conjunction of the supervenience requirement with the main claim of the knowledge-first approach suggests a *prima facie* incompatibility: if, by **(ES)**, epistemic properties are required to supervene on non-epistemic base, knowledge firsters seem to deny this – arguably, one claim implicitly assumed by knowledge firsters is that relevant epistemic properties supervene on knowledge, usually considered an epistemic property *par excellence*.<sup>7</sup> We can better outline the base properties problem as follows:

$P_1$  By epistemic supervenience **(ES)**, necessarily, every epistemic property (necessarily) supervenes on a non-epistemic property.

$P_2$  In the knowledge-first approach, relevant epistemic properties (*justification, evidence*) supervene on a fundamental epistemic property, i.e. *knowledge*, rather than on a non-epistemic property.<sup>8</sup>

C: Knowledge-first epistemology violates **(ES)**.

Let's begin by noticing that the proponents of the objection need to make an important assumption to press the base properties problem, namely that supervenience is *exclusive*. Crudely put, the assumption is the following: if A supervenes on B, then A cannot supervene also on C. However, it should be stressed that supervenience is not exclusive as the proponents of the objection take it to be. It is in fact compatible with supervenience that epistemic properties supervene on knowledge and *also* on other non-epistemic properties. Indeed, this is to be expected given the transitivity of supervenience. In fairness to the proponents of the objection, knowledge-firsters nevertheless violate **(ES)** even if supervenience isn't exclusive, since relevant epistemic properties supervene on an epistemic property, i.e. *knowledge*. Absent any clarification of the further non-epistemic base upon which knowledge supervenes, the objection still stands. Be that as it may, pressing the base properties problem against knowledge firsters is still instructive, because it bids knowledge firsters to identify the non-epistemic base of these relevant epistemic properties.

---

<sup>7</sup>Here is Turri: “Epistemic properties include *justification, rationality, reasonableness, warrant, and knowledge*” (Turri 2010: 1).

<sup>8</sup>The reader might worry that friends of the knowledge-first approach are not committed to this supervenience claim. Admittedly, knowledge-first epistemology does not *imply* any such claim. However, this supervenience claim plausibly underlies and captures the explanatory priority of knowledge over justification or evidence, as defended by Ichikawa (2014: 185), Miracchi (2015: §4.3) and Kelp (2017: 66-67).

Coming back to the objection: fortunately enough for the advocates of the knowledge-first approach,  $P_2$  is false. To see why, recall that in a knowledge-first setting, knowledge is not *merely* an epistemic property; rather, knowledge is a factive *mental state* in its own right (Williamson 2000: chapter 1). As such, it qualifies as subvenient *mental* base for epistemic properties such *justification* and *evidence*. The reasoning underlying the base properties problem enjoys prima facie plausibility, but a closer inspection reveals it to be ultimately misleading: this strand of the objection does not cause any concern for knowledge firsters who hold that knowledge is a factive mental state.<sup>9</sup> Accordingly, here is a plausible counterargument to the base properties problem:

- $P_1$  By epistemic supervenience (**ES**), necessarily, every epistemic property supervenes on a non-epistemic property.
- $P_2$  In the knowledge-first approach, relevant epistemic properties (*justification*, *evidence*) supervene on a *mental state*, i.e. *knowing*.
- C: Knowledge-first epistemology accommodates (**ES**): epistemic properties supervene on a non-epistemic *mental* base.

$P_2$  can be outlined more clearly as follows:

**Knowledge-first Epistemic Supervenience.** Necessarily, relevant epistemic properties necessarily supervene on a subvenient non-epistemic base, i.e. a (factive) mental state (knowing).<sup>10</sup>

It is worth making two important points of clarification. Firstly, I should emphasise that the view that knowledge is a mental state has received several criticisms.<sup>11</sup> However, these criticisms need not concern us here: this argument does not support the view that knowledge is a mental state. Rather, it aims for a weaker conditional claim: *if* knowledge is a mental state, then the base properties problem is easily addressed. Secondly, the reader might be puzzled by the two sets of properties discussed here – epistemic properties and mental properties of mental states. But this is certainly no news: evidentialist mentalists such as Conee and Feldman (2004: Chapter 3) also hold that justifiers are mental states.<sup>12</sup> Knowledge firsters apply the same evidentialist mentalist template, but with a

<sup>9</sup>The proud few include Bird (2007), Nagel (2013) and Miracchi (2015). The view is endorsed also by authors who do not take themselves to be contributing to the knowledge-first programme, such as Schellenberg (2018: §2) and Sylvan (2018).

<sup>10</sup>A word of caution. Some contributors to the knowledge-first programme will probably reject this formulation in terms of supervenience, and advocate for a different *hyperintensional* relation holding between the epistemic and non-epistemic domains. For instance, Ichikawa (2014: 185) formulates his knowledge-first project in terms of (relative) metaphysical priority rather than supervenience. In this appendix, I adopt the supervenience formulation for the following three reasons. The first reason has to do with my aim to *charitably* reconstruct the supervenience objection: accordingly, I borrow the proponents' formulation in terms of supervenience. Secondly, supervenience is a coarse-grained metaphysical relation (see also section 4 of this appendix, fn 22), broad enough to encompass also views like Ichikawa's. Third, the proponents of the objection could run the same type of argument against knowledge firsters by using a different metaphysical relation (e.g., metaphysical priority).

<sup>11</sup>McGlynn (2012: 167-195), Fricker (2009), Smith (2017).

<sup>12</sup>A quick but important note on the terminology adopted here. I use the term "mentalism" to refer to any view that unpacks epistemic properties in terms of mental states. Mentalism is nevertheless silent on

crucial difference in the type of mental states that confer justification: while Conee and Feldman hold that non-doxastic, internally accessible experiences confer justificatory status to beliefs, knowledge firsters typically hold that only factive, truth-entailing mental states justify beliefs (e.g., Williamson 2000 and Millar 2014).

Before moving on to the second strand of the objection, let's consider a possible rejoinder on behalf of the knowledge-first critics. As they correctly observe (Pritchard and Kallestrup 2017: §1), one of the main motivations for **(ES)** is *physicalism*: in virtue of mild physicalist commitments, all (positive, non-indexical) properties of our world strongly supervene on physical properties of our world, where a property is physical just in case it occurs in current physical theory (or an improved version thereof). Therefore, it follows that mental states (knowledge included) are equally subject to a supervenience constraint. If that is the case, what is the non-mental, physical base of this particular type of mental state, i.e. knowledge?

To make progress on this question, a brief detour in the metaphysics of normativity is needed. Recall that epistemic properties such as *justification*, *evidence* and *rationality* are a subset of a broader class of normative properties – *right*, *good*, *virtuous* etc etc. It is in fact not surprising that epistemology is widely taken to be a normative discipline, and attempts to unify ethics and epistemology are becoming increasingly common in the literature.<sup>13</sup> The challenge faced by knowledge-firsters is to explain epistemic normative properties in terms of a paradigmatic epistemic normative property, i.e. *knowledge*. Nonetheless, there is room for doubt that knowledge has a normative nature just as other epistemic properties seem to have. Borrowing useful terminology from Kurt Sylvan (2018: §1), I dub the view that sees knowledge as having a wholly normative nature *Epistemic Normativism*. With Epistemic Normativism in play, let's proceed to see how it can help to address this potential rejoinder.

Epistemic Normativism is best understood as a metaphysical view on the constitution of normative properties: following Sylvan (2018: 192-193), we can say that a property is normative if and only if it is identical or partly constituted by a normative item. Back to epistemology: many of the post-Gettier analyses of knowledge mention justification, a paradigmatic normative property, as *analysans*.<sup>14</sup> Given the metaphysical account of normativity in play here, these analyses are all committed to Epistemic Normativism: knowledge has a normative nature inherited by its normative constituents (e.g., justification). By contrast, the knowledge-first approach goes against it in at least two ways. Firstly, it denies that justification is a constituent of knowledge, hence denying that knowledge has a paradigmatically normative constituent such as justification. Secondly, if knowledge is

---

the relevant type of mental state in question. I should also warn the reader that Pritchard (2012b: 238) understands mentalism as the view that assigns equal justification to two mentally identical subjects. I am not following Pritchard in my choice of terminology.

<sup>13</sup>This is certainly no news: Clifford's deservedly influential work might be taken to show that epistemology is up to its neck in ethics (Clifford 1877). However, also more contemporary work in virtue epistemology (Zagzebski 1996, Sosa 2015) proves the same point.

<sup>14</sup>Notable exceptions here include Goldman's early process-reliabilism (Goldman 1976) and Foley's divorce of a theory of knowledge from a theory of justified belief (Foley 1987). Sylvan (2018: 196) also quotes these two relevant examples.

the most general factive mental state, then knowledge is not constituted by any normative element. Mental states are not normative in nature, even though they might have normative significance. For instance, belief is a mental state with specific normative standards, but these normative standards do not make it normative (compare: chess also has normative constraints, but this does not make the game of chess normative in nature). To make this point clearer, it is helpful to abandon talk of metaphysical constitution in favour of a more fine-grained relation – the determinable/determinate relation. Knowledge is the most general determinable of a specific set of determinates (factive mental states such as *seeing that p*, *hearing that p*, *perceiving that p*...) that entail knowledge. Williamson does not say more on the nature of such mental states, but we can make two important observations. If the determinates are not normatively constituted, then the determinable is not either.<sup>15</sup> Second, the determinates of knowledge are not normatively constituted: they have a non-normative nature just like any other mental state (whether factive or not). Accordingly, it follows that knowledge lacks a normative nature.

Building on Lyons (2009: 88-98) and Sylvan (2018: 197-198), one way to characterise the physical base of the determinates of knowledge is to appeal to modules and cognitive systems. In a nutshell, factive mental states amount to modular systems that give access to a fact and store it in the central system, as Fodor would put it.<sup>16</sup> Intuitively, modules are *not* normative in nature – they are sub-personal cognitive mechanisms that ground personal level mental states. As sub-personal cognitive mechanisms they have a physical base, and such base constitutes the relevant physical base of factive mental states.<sup>17</sup> Thus, given the non-normative nature of its determinates, knowledge is non-normative too. And if knowledge is not normative, then it properly qualifies as the non-normative mental base for other epistemic properties. Perhaps unsurprisingly, views labelled as mentalism can satisfy the epistemic supervenience requirement.<sup>18</sup>

It might be argued that despite being a mental state, knowledge nevertheless retains a fundamentally epistemic dimension. Perhaps it has a dual nature, both mental and epistemic.<sup>19</sup> If so, the epistemic constituent would prevent it from qualifying as the suitable non-epistemic base I am taking it to be. This is a fair worry, that I hope to soothe as follows. Viewing knowledge as a mental state amounts to treat knowledge as a particularly significant and non-normative perceptual relation – the relation one is in if one is in a factive relation at all. Relations to facts are epistemically significant, yet they need not be epistemic in nature. Belief is another case in point. Let's take belief as a relation

---

<sup>15</sup>Worley (1997) makes this very same point with respect to mental causation, as Sylvan (2018) observes.

<sup>16</sup>Post Fodorian theorists such as Carruthers (2006) will appeal to a different notion of modularity to unpack factive mental states, while other epistemologists will wholly reject the notion (Lyons 2009: chapter 4). Here, I adopt Fodor's notion of modularity to keep the discussion manageable, but the reader is free to adopt the reading she prefers.

<sup>17</sup>Lisa Miracchi (2015: 46) also unpacks factive mental states in terms of sub-personal cognitive mechanisms. Unlike Sylvan and Lyons, she explicitly identifies her project as a contribution to the knowledge-first research programme. As such, her proposal qualifies as an actual example of knowledge-first epistemology which escapes the base properties problem.

<sup>18</sup>Turri notices this feature of mentalism in his discussion of epistemic supervenience (Turri 2010: 8).

<sup>19</sup>See French (2012) for a thorough discussion of these two components and their relationship.

to a proposition: relations to propositions are undoubtedly epistemically significant, yet they need not be epistemic in nature either. Knowledge is then an epistemically privileged perceptual relation, but, qua perceptual relation, it hardly implies epistemic or normative constituents.

Let's take stock now. By epistemic supervenience, epistemic properties supervene on a non-epistemic base. Given that knowledge is a factive mental state, it properly qualifies as such base. Nevertheless, mental states are equally subject to a similar supervenience requirement. But unpacking personal level mental states in terms of sub-personal cognitive mechanisms addresses also these further physicalist concerns.<sup>20</sup> Thus, *pace* the proponents of the supervenience objection, the first strand of the objection does not go through.

## The Epistemic Superdupervenience Demand

In order to fully appreciate the challenge raised by the epistemic superdupervenience demand we need to take another brief detour to understand the role played by superdupervenience in metaphysics. It will be helpful to start by noticing how metaphysicians have lately started to question modal notions such as supervenience, expressing doubts concerning its explanatory power.<sup>21</sup> These perplexities are not new though: Jaegwon Kim (1993: 167) voiced them nicely in his discussion of supervenience:

Supervenience itself is not an explanatory relation. It is not a “deep” metaphysical relation; rather, it is a “surface” relation that reports a pattern of property covariation, suggesting the presence of an interesting dependency relation that might explain it.

The proponents of the supervenience objection share Kim's worries, and in addition to the individuation of a proper non-epistemic base, they also require a more fine-grained metaphysical relation that explains the “surface” supervenience relation.<sup>22</sup> This relation goes

---

<sup>20</sup>Here is an objection that was raised against my treatment of the base properties problem.

*Objection:* doesn't the appeal to sub-personal cognitive mechanisms collapse knowledge-first epistemology into process reliabilism? If a full knowledge-first explanation of epistemic properties ultimately appeals to sub-personal cognitive mechanisms, aren't we back to traditional process reliabilism?

*Answer:* No. Appealing to sub-personal cognitive mechanisms sheds light on the nature of factive mental states, but it does not *explain* evidence and justification. Knowledge firsters can concede that factive mental states supervene on (or are grounded in) a physical base constituted by sub-personal cognitive mechanisms *without* explaining epistemic properties in terms of such base. Justification and evidence will be explained in terms of knowledge, rather than in terms of the physical base of knowledge. Notice further that knowledge-firsters need not place a reliability condition on knowledge and cognate notions (although some do: see for instance Miracchi 2015, Kelp 2017, Simion 2019). Thus, there is no risk of collapse. Section 5 of this appendix and fn 32 address for similar concerns.

<sup>21</sup>See Fine (1994) and Wilson (2014: §1) for discussion.

<sup>22</sup>The reader might wonder about the difference between “fine-grained” and “coarse-grained” metaphysical relations. To a rough approximation, while the latter are usually deemed to be too general (and therefore compatible with a plethora of other metaphysical relations), the former indicate a specific way of dependence. Supervenience is almost universally taken to be a coarse-grained relation, whilst identity, functional realization and the determinable-determinate relation are prime examples of fine-grained relations. See Wilson (2014) and Bennett (2017) for discussion.

under the name of “superdupervenience”, and it’s supposed to explain how exactly epistemic properties supervene on their non-epistemic base – which has now been qualified as constituted by mental states. Crucially, the proponents of the supervenience objection allege that knowledge firsters have not provided (and cannot provide) this important fine-grained explanatory relation (Pritchard and Kallestrup 2017: 18). Hence, by exploiting this further explanatory demand stemming from supervenience, they use also the superdupervenience requirement to question the metaphysical viability of knowledge-first epistemology.

In order to deal with the second part of the objection, I shall distinguish two different lines of response: a refusal and an acceptance line. The main difference between the two lines lies in whether the superdupervenience demand is ultimately legitimate. In what follows, I fully develop these two lines, and count costs and benefits of each.

## The Refusal Line

One initial reaction to the superdupervenience demand is to simply deny that the requirement is fair. First off, the kind of fine-grained metaphysical relations that provide a further explanation of supervenience are not common in epistemology: different advocates of different views are happy to formulate their position without going into such detail, and while some authors can be read as providing an implicit superdupervenience in reductionist or functionalist terms, very few authors have provided such superdupervenience explicitly. For instance, implicit versions of superdupervenience in epistemology plausibly include Goldman (1979), who offers a reductive, identity-based account of justification or Plantinga (1993), Graham (2014b) and Simion (2019), who explain epistemic warrant and justification in terms of functional realization. But again, none of them explicitly provides an explanation of the supervenience relation holding among these properties.<sup>23</sup> To wit, let’s focus again on evidentialism mentalism: Conee and Feldman (2004: 101) famously argued that evidential facts supervene on a specific set of mental facts – occurrent or dispositional non-doxastic mental states that support the truth of a given proposition. Their view has attracted a number of criticisms, yet nobody has ever pressed Conee and Feldman to explain how evidential facts superdupervene on mental facts, or why exactly this supervenience relation holds. Crudely put: if evidentialists have not been pressed on this particular matter, why should knowledge firsters be pressed too?

Secondly, knowledge firsters can point out that the superdupervenience demand is unfair because it implicitly assumes a default commitment to reductionist views, without leaving any non-reductionist explanation open. Here, “reduction” means to posit a relation of *identity* between epistemic and non-epistemic properties. In order to make this commitment clearer, it is worth considering an analogous case in metaethics. Moral properties, which also fall under the category of evaluative and normative properties, are equally sub-

---

<sup>23</sup>One notable exception is Zangwill (2018), who develops a notion of epistemic dependence by appealing to fine-grained metaphysical relations (such as metaphysical grounding and ontological dependence).

ject to an identical supervenience (and superdupervenience) constraint.<sup>24</sup> While there's no consensus about what exactly such superdupervenience consists of, this much is agreed upon: reductive explanations that posit an identity relation between moral and natural properties have a very easy time explaining the modal co-variation holding between these two sets of properties. Trivially, if moral properties are identical with natural properties, then obviously there can't be a change in the former without a change in the latter.<sup>25</sup> In view of this, knowledge firsters can reject the superdupervenience demand because reductionism seems to be the only agreed way to heed it.

Given its simplicity, there's something undoubtedly attractive about this way of dealing with the superdupervenience demand; nevertheless, it carries a great cost. In fact, to fully pursue this line leads to a pernicious dialectical stalemate: on the one hand, the proponents of the supervenience objection demand a further explanation of supervenience by leaning on seemingly uncontentious constraints commonly accepted in metaphysics; on the other, knowledge firsters point out that such demand is unfair by making equally plausible observations. Both parties stick to their guns, and progress is hardly made.

But, not all is lost. While the superdupervenience demand may not be totally fair, it is certainly more charitable for knowledge firsters to compromise with their opponents, and (at least attempt to) meet their demand. Luckily enough, this can be done by providing two different types of superdupervenience – a reductive superdupervenience in terms of *identity* and a non-reductive superdupervenience in terms of *metaphysical grounding*.

## The Acceptance Line

### First Stab: Reduction and Identity

Alternatively, knowledge firsters may take up also the challenge underlying the superdupervenience demand and provide a suitable, metaphysical fine-grained relation that explains why supervenience holds in the first place. Building on the points made in the previous section, the reductionist line is certainly available to knowledge firsters. All is needed is an identity relation holding between epistemic facts (evidential facts, justification facts) and mental facts. Accordingly:

**Knowledge-first Reductionism.** Necessarily, if subject S has *justification* to believe *p* and/or has *evidence* supporting *p*, then these epistemic facts are identical with mental facts (e.g., the fact that S knows *p*).<sup>26</sup>

Intuitively enough, a relation of identity easily explains the pattern of modal co-variation suggested by supervenience, and this is an obvious benefit. However, such benefit comes

---

<sup>24</sup>See McPherson (2022: §4) for relevant discussion.

<sup>25</sup>Cornell realists all embrace some version of moral naturalism. The *locus classicus* for Cornell realism is Boyd (1988). Jackson (1998) is also an advocate of moral naturalism, albeit in a more reductionist version. See Miller (2013: chapter 8 and 9) for discussion of reductionist and non-reductionist versions of moral naturalism.

<sup>26</sup>For simplicity sake, this formulation targets *doxastic* justification.

with a price. Notice that the move is available only to *strong* knowledge-first reductive views – views that identify justified belief with knowledge.<sup>27</sup> Unfortunately, these views are strongly revisionary and notoriously face a number of objections on independent grounds. A popular complaint against knowledge-first reductionism is that it simply can't accommodate important traditional roles that the notion of justification is supposed to play. First, while justification intuitively comes in degrees, knowledge is an all-or-nothing matter. Justification ascriptions often take the form of comparative judgements (we deem certain beliefs to be more or less justified than others), but propositional knowledge does not allow for such comparative evaluations. So, if justification just *is* knowledge, how is the *gradability* of justification to be explained?<sup>28</sup> Secondly, by identifying knowledge and justification champions of strong knowledge-first reductive views can hardly account for intuitively different types of cases.<sup>29</sup> To wit, justification is denied to both true beliefs of intuitively justified victims of Gettier cases and also to epistemically responsible false beliefs formed by radically deceived subjects. Some false beliefs can be justified even though they fall short of knowledge: this is a key distinction worth marking in epistemology, and strong knowledge-first reductive views do not seem to be well-positioned to mark it.<sup>30</sup> Hence, given how revisionary these views are, the costs of an identity-based explanation of supervenience outweigh the benefits.

## Doing Better: Non-reduction and Metaphysical Grounding

A more promising variation of the acceptance line consists in appealing to a different fine-grained metaphysical relation: *metaphysical grounding*. These days, grounding is all the rage in metaphysics.<sup>31</sup> More precisely, metaphysical grounding amounts to “an objective and mind-independently obtaining hyperintensional and non-monotonic strict partial ordering relation, asymmetric, irreflexive, and transitive, always holding between what is less and what is more fundamental such that the more fundamental grounds the less fundamental, but not vice versa” (Maurin 2019: 1). This formulation captures the orthodox view of metaphysical grounding: in fact, while grounding theorists have internal disagreements concerning different features of grounding, they all agree on one essential point: grounding relations are explanatory – they provide a distinct form of metaphysical explanation. Since the superdupervenience demand is roughly an explanatory challenge, metaphysical grounding is a natural candidate to explain why supervenience relations hold: epistemic properties supervene on the mental because they are metaphysically grounded in these more fundamental knowledge states. More explicitly:

---

<sup>27</sup>E.g., Sutton (2007); Littlejohn (forthcomingb), Williamson (forthcoming).

<sup>28</sup>McGlynn (2014: 41, 42) makes this point, and Brown (forthcoming: chapter 4) also expands on this.

<sup>29</sup>See McGlynn (2014: Chapter 3) and Kelp (2016: §1) for relevant discussion.

<sup>30</sup>In fairness, advocates of strong knowledge-first reductive views have developed a system of primary, secondary and tertiary norms and excuses to deal with these cases (Williamson, forthcoming; Littlejohn forthcomingb).

<sup>31</sup>Bliss and Trogon (2021), Correia and Schnieder (2012).

**Knowledge-first Metaphysical Grounding.** Necessarily, if subject S has *justification* to believe *p* and/or has *evidence* supporting *p*, then these epistemic facts are grounded in/metaphysically explained by facts about S's mental states (e.g., the fact that S knows *p*).<sup>32</sup>

Before counting the costs and the benefits of this variation of the acceptance line, it is worth making an important point of clarification. I'm not the first one to extend the grounding research programme to epistemology: recently, epistemologists have started to look at metaphysical grounding relations with increasing interest (Beddor 2015, Elgin 2018). However, I care to emphasise that I have *no stake* at defending metaphysical grounding *per se* here: even if it's not the kind of deeper metaphysical relation that Kim originally envisaged, it can serve as the sort of superdupervenience requested by the proponents of the objection.<sup>33</sup>

Since epistemic facts and mental facts are distinct, metaphysical grounding yields a *prima facie* non-reductive version of superdupervenience. Thus, it is available to a larger number of knowledge firsters without a default commitment to strong reductionist versions of the knowledge-first approach. Accordingly, it carries a valuable benefit that an identity-based explanation of supervenience lacks. However, one might wonder whether grounding and identity are actually two different metaphysical relations, and admittedly some grounding theorists disagree on the matter.<sup>34</sup> Still, an initial and seemingly plausible case can be made for divorcing grounding and identity by insisting that the two exhibit different formal properties. Simply put, identity, unlike grounding, is symmetric and reflexive. Accordingly, grounding and identity come apart, and a superdupervenience in terms of metaphysical grounding qualifies as the most beneficial strategy that knowledge firsters can pursue.

Let's take stock now. In sections 5 and 6, I laid out two different ways to deal with the epistemic superdupervenience demand. While each of these lines is worth pursuing for different reasons and depending on one's theoretical commitments, I submit that the acceptance line in terms of metaphysical grounding fares better. Be that as it may, also this second strand of the supervenience objection clearly fails.

## A Final Dilemma: Knowledge First?

The last strand of the supervenience objection does not have to do directly with supervenience constraints but exploits an intuitive consequence stemming from them. The risk

---

<sup>32</sup>See Beddor (2015: §7) for a similar formulation. Notice also that this would be a case of *partial* grounding, since mental facts will in turn be grounded in more fundamental physical facts. As we've seen above, it's plausible that the fact that S has a personal level mental state is grounded in more basic physical facts involving the proper functioning of S's sub-personal cognitive mechanisms. For the purposes of explaining supervenience relations, this shouldn't matter. All that matters is that metaphysical grounding relations hold between epistemic and mental facts.

<sup>33</sup>In passing, they concede that metaphysical grounding can count as a satisfactory superdupervenience (Pritchard and Kallestrup 2017: 18).

<sup>34</sup>See Rosen (2010) and Wilson (2014) for relevant discussion.

here is the following: even if knowledge firsters can meet the supervenience challenge by identifying a non-epistemic base and by explaining why supervenience holds, epistemic supervenience ends up being essentially incompatible with putting knowledge first. The proponents of the objection argue as follows:

- $P_1$  By knowledge-first epistemic supervenience, necessarily, epistemic properties supervene on a factive mental state, knowledge, which in turn supervenes on a more fundamental physical base.
- $P_2$  By transitivity of supervenience, relevant epistemic properties *ultimately* supervene on the more fundamental physical base on which knowledge supervenes.
- $P_3$  In the knowledge-first approach, “first” means “*metaphysically prior*”: knowledge should have a *relevant* metaphysical priority over other epistemic properties. Such relevant priority simply goes missing given that, by transitivity of supervenience, epistemic properties do not ultimately supervene on knowledge.
- C:** Meeting the supervenience constraints is incompatible with putting knowledge first. Alternatively, in the form of a dilemma: either knowledge is put first, or epistemic supervenience is met.

Admittedly,  $P_3$  is not made explicit in the discussion of the objection, but it is strongly hinted:

Reflect finally that even if Williamson were able to accommodate **(ES)**, a second problem would appear. Suppose that evidence and justification supervene on non-epistemic properties, because these epistemic properties supervene on knowledge which in turn supervenes on non-epistemic properties. Remember that supervenience is transitive. *And the question is of course whether knowledge is still being put first if such an account would provide a supervenience base for evidence and justification in terms of wholly non-epistemic properties.* (Pritchard and Kallestrup 2017: 17, 18; emphasis added)

There is nevertheless room for blocking the conclusion of the last strand of the objection, and showing that it amounts to a false dilemma. In fact, knowledge firsters have different options to vindicate the priority of knowledge and, at the same time, meeting the supervenience requirement. In the remainder of this section, I will outline two of these options and explain how they both address the last strand of the objection.

A first promising line of response assumes a more precise gloss of ‘priority’ as starting point. Ichikawa (2014) argues that knowledge has a *relative metaphysical priority*: knowledge is prior to epistemic properties such as *evidence* or *justification*, but not to more fundamental physical states.<sup>35</sup> After all, knowledge-first epistemology should be compatible with an in principle reduction of knowledge to more basic physical states, as physicalism predicts. Here Ichikawa expressing such point nicely:

---

<sup>35</sup>See Sider (2011) for the notion of relative priority.

As I understand it, the claim that knowledge is “first” should be read as a claim about *relative priority*; knowledge is prior to the salient alternate states that one might be interested in. I shall suggest that knowledge is metaphysically prior to justification. There is no commitment to the stronger claim that there is nothing prior to knowledge. (...) The knowledge-first stance is consistent with, e.g. physicalism, according to which atoms are prior to knowledge. (Ichikawa 2014:185)

Alternatively, one could appeal to distinct kinds of priorities of knowledge, and argue that although a certain kind of metaphysical priority may very well fail, at the conceptual level knowledge does indeed come first. Carrie Jenkins and Jonathan Ichikawa, in joint work, propose a taxonomy of priorities of knowledge, usefully distinguishing between metaphysical and representational priorities: the former involves properties and extensions, whilst the latter is a *cognitive priority* located at the level of thought and language (Ichikawa and Jenkins 2017: §2). Moreover, even though the two different kinds of priorities are not incompatible, they are nevertheless independent, to the extent that one does not entail the other.

Accordingly, knowledge firsters can exploit such different priorities and argue as follows: while the *concept* of knowledge cannot be reductively analysed in terms of individually necessary and jointly sufficient conditions, its *extension* smoothly supervenes on more basic (physical) states. Simply put, the priority of a concept need not (and often does not) go with the priority of its referent. Here is how to better flesh out this strategy. Sosa (1991: 153-154; 2015: 8-10), while discussing the ‘doctrine of epistemic supervenience’, dubs such view that divorces concepts and extensions as ‘pessimism’. Interestingly, by adapting Sosa’s pessimism to their framework, knowledge firsters can easily vindicate their anti-reductionist stance on knowledge by denying that knowledge is susceptible of a reductive analysis, and nevertheless conceding, with physicalists and the proponents of the supervenience objection, that worldly instances of knowledge supervene on a suitable non-epistemic base.<sup>36</sup> It is also worth noticing that the kind of pessimism described here nicely dovetails with conceptual atomism. The view has been famously defended by Fodor (1998) and Millikan (1993): roughly, conceptual atomists hold that lexical concepts have no semantic structure, and therefore cannot be reductively analysed in terms of individually necessary and jointly sufficient conditions. Given that *knows* is a lexical concept, the view correctly predicts that it can’t be analysed.<sup>37</sup>

Here what the two options consist of. While the second best captures different priorities of knowledge within the knowledge-first approach, both successfully address the last remaining part of the supervenience objection.

---

<sup>36</sup>Compare: advocates of the phenomenal concept strategy in philosophy of mind also hold that mental concepts (such as phenomenal concepts) cannot be reductively analyzed, and yet they take these concepts to refer to properties that supervene on non-mental properties. See Stoljar (2005) for discussion.

<sup>37</sup>Perhaps unsurprisingly, Williamson (2000: 31, fn 4) briefly mentions Fodor’s theory of concepts as relevant, and so do also Ichikawa and Jenkins (2017: 17).

## Concluding remarks

In this appendix, firstly I charitably reconstructed the supervenience objection levelled against knowledge-first epistemology, and then I showed how knowledge firsters can successfully address it. The objection is three-fold, but each strand ultimately fails. The base properties problem is solved by leaning on one of knowledge-first epistemology central tenets, i.e. that knowledge is a mental state. Similarly, the epistemic superdupervenience demand is also easily dealt with. I outlined two different ways to defuse this second strand of the objection: a refusal line, which questions the legitimacy of the demand, and an acceptance line which consists in explaining supervenience in terms of identity and metaphysical grounding. While each line carries its own costs and benefits, they both satisfy the superdupervenience demand. Finally, I offered two strategies to address the last part of the objection, dubbed as the failure of priority. The strategy with more far-reaching potential consists in ascribing different priorities to knowledge, representational and metaphysical, and then concede that while knowledge itself may not be metaphysically fundamental, its concept cannot be reductively analysed. It follows that the supervenience objection ultimately does not stick. Thus, it is fair to conclude that knowledge-first epistemology has strong metaphysical credentials, even stronger than its proponents have hitherto appreciated.

# Bibliography

- Mark Alfano. Sensitivity theory and the individuation of belief-formation methods. *Erkenntnis*, 70(2):271–281, 2009. doi: 10.1007/s10670-008-9127-9.
- Ronald Allen. Naturalized epistemology and the law of evidence revisited. *Quaestio facti. Revista internacional sobre razonamiento probatorio*, 0(2), 2021a. ISSN 2604-6202. doi: 10.33115/udg\_bib/qf.i2.22446. URL <https://revistes.udg.edu/quaestio-facti/article/view/22446>.
- Ronald Allen. Naturalized epistemology and the law of evidence revisited: Reply to david enoch. *Quaestio facti. Revista internacional sobre razonamiento probatorio*, 0(2), 2021b. ISSN 2604-6202. doi: 10.33115/udg\_bib/qf.i2.22446. URL <https://revistes.udg.edu/quaestio-facti/article/view/22446>.
- Ronald Allen. Naturalized epistemology and the law of evidence revisited: Reply to pardo. *Quaestio facti. Revista internacional sobre razonamiento probatorio*, 0(2), 2021c. ISSN 2604-6202. doi: 10.33115/udg\_bib/qf.i2.22446. URL <https://revistes.udg.edu/quaestio-facti/article/view/22446>.
- Ronald J. Allen. Explanationism all the way down. *Episteme*, 5(3):320–328, 2008.
- Ronald J. Allen and Brian Leiter. Naturalized epistemology and the law of evidence. *Virginia Law Review*, 87(8):1491–1550, 2001. ISSN 00426601. URL <http://www.jstor.org/stable/1073901>.
- Ronald J. Allen and Michael S. Pardo. Relative plausibility and its critics. *The international journal of evidence & proof*, 23(1-2):5–59, 2019.
- Ronald J. Allen and Alex Stein. Evidence, probability, and the burden of proof. *Arizona law review*, 55(3):557, 2013.
- Maria Rosa Antognazza. The benefit to philosophy of the study of its history. *British Journal for the History of Philosophy*, 23(1):161–184, 2015. doi: 10.1080/09608788.2014.974020.
- Marvin Backes. Epistemology and the law: Why there is no epistemic mileage in legal cases. *Philosophical Studies*, 177(9):2759–2778, 2020. doi: 10.1007/s11098-019-01337-6.

- Sam Baron, Mark Colyvan, and David Ripley. How mathematics can make a difference. *Philosophers' Imprint*, 17, 2017.
- C. Van Fraassen Bas. *The Scientific Image*. Oxford, England: Oxford University Press, 1980.
- Peter Baumann. Is knowledge safe? *American Philosophical Quarterly*, 45(1):19–30, 2008.
- Kelly Becker. *Methods and how to individuate them*, page 81–98. Cambridge University Press, 2012. doi: 10.1017/CBO9780511783630.008.
- Kelly Becker. *The Sensitivity Response to the Gettier Problem*, page 108–124. Classic Philosophical Arguments. Cambridge University Press, 2018. doi: 10.1017/9781316827413.009.
- Bob Beddor. Evidentialism, circularity, and grounding. *Philosophical Studies*, 172(7): 1847–1868, 2015. doi: 10.1007/s11098-014-0375-z.
- Bob Beddor and Carlotta Pavese. Modal virtue epistemology. *Philosophy and Phenomenological Research*, 101(1):61–79, 2018. doi: 10.1111/phpr.12562.
- Marcello Di Bello. Trial by statistics: Is a high probability of guilt enough to convict? *Mind*, 128(512):1045–1084, 2019. doi: 10.1093/mind/fzy026.
- Karen Bennett. *Making Things Up*. Oxford University Press, 2017.
- Michael Bergmann. *Justification Without Awareness: A Defense of Epistemic Externalism*. Oxford University Press, 2006.
- Sven Bernecker. Sensitivity, safety, and closure. *Acta Analytica*, 27(4):367–381, 2012. doi: 10.1007/s12136-011-0137-x.
- Sven Bernecker. Against global method safety. *Synthese*, 197(12):5101–5116, 2020. doi: 10.1007/s11229-018-02008-5.
- Sven Bernecker. *An Explanationist Account of Knowledge*. forthcoming.
- Francesco Berto and Daniel Nolan. Hyperintensionality. In *Stanford Encyclopedia of Philosophy*. 2021.
- Alexander Bird. Justified judging. *Philosophy and Phenomenological Research*, 74(1): 81–110, 2007. doi: 10.1111/j.1933-1592.2007.00004.x.
- Ricki Bliss and Kelly Trogon. Metaphysical Grounding. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.

- Michael Blome-Tillmann. Sensitivity, causality, and statistical evidence in courts of law. *Thought: A Journal of Philosophy*, 4(2):102–112, 2015. doi: 10.1002/tht3.163.
- Michael Blome-Tillmann. “More Likely Than Not” - knowledge first and the role of statistical evidence in courts of law. In Adam Carter, Emma Gordon, and Benjamin Jarvis, editors, *Knowledge First - Approaches in Epistemology and Mind*, pages 278–292. Oxford, UK: Oxford University Press, 2017.
- Michael Blome-Tillmann. Non-reductive safety. *Belgrade Philosophical Annual*, 33:25–38, 2020.
- Tomas Bogardus. Knowledge under threat. *Philosophy and Phenomenological Research*, 88(2):289–313, 2014. doi: 10.1111/j.1933-1592.2011.00564.x.
- Tomas Bogardus and Will Perrin. Knowledge is believing something because it’s true. *Episteme*, pages 1–19, forthcoming. doi: 10.1017/epi.2020.18.
- Laurence Bonjour. Externalist theories of empirical knowledge. *Midwest Studies in Philosophy*, 5(1):53–73, 1980. doi: 10.1111/j.1475-4975.1980.tb00396.x.
- Richard Boyd. How to be a moral realist. In G. Sayre-McCord, editor, *Essays on Moral Realism*, pages 181–228. Cornell University Press, 1988.
- Fernando Broncano-Berrocal and J. Adam Carter. Epistemic luck. In Tim Crane, editor, *Routledge Encyclopedia of Philosophy*. Routledge, 2017.
- Jessica Brown. Vi-reliabilism, knowledge, and mental content. *Proceedings of the Aristotelian Society*, 100(2):115–135, 2000. doi: 10.1111/1467-9264.00070.
- Jessica Brown. Precis of fallibilism: Evidence and knowledge. *Philosophical Studies*, pages 1–4, forthcoming. doi: 10.1007/s11098-022-01781-x.
- Lara Buchak. Belief, credence, and norms. *Philosophical Studies*, 169(2):1–27, 2014. doi: 10.1007/s11098-013-0182-y.
- David J. Buller. Etiological theories of function: A geographical survey. *Biology and Philosophy*, 13(4):505–527, 1998. doi: 10.1023/A:1006559512367.
- Guido Calabresi. Some thoughts on risk distribution and the law of torts. *The Yale Law Journal*, 70(4):499–553, 1961. ISSN 00440094. URL <http://www.jstor.org/stable/794261>.
- Peter Carruthers. *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought: Massive Modularity and the Flexibility of Thought*. Oxford University Press UK, 2006.
- J. Adam Carter. Robust virtue epistemology as anti-luck epistemology: A new solution. *Pacific Philosophical Quarterly*, 97(1):140–155, 2016. doi: 10.1111/papq.12040.

- J. Adam Carter. De minimis normativism: A new theory of full aptness. *Philosophical Quarterly*, 71(1):16–36, 2021. doi: 10.1093/pq/pqaa017.
- J. Adam Carter. *Bi-Level Virtue Epistemology: A Defence*. Cambridge University Press, forthcoming.
- J. Adam Carter and Ernest Sosa. Metaepistemology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2022 edition, 2022.
- J. Adam Carter, Emma C. Gordon, and Benjamin W. Jarvis. *Knowledge First: Approaches in Epistemology and Mind*. Oxford: Oxford University Press, 2017.
- Quassim Cassam. Can the concept of knowledge be analysed? In Patrick Greenough and Duncan Pritchard, editors, *Williamson on Knowledge*. Oxford University Press, 2009.
- Roderick Milton Chisholm. *Theory of Knowledge*. Englewood Cliffs, NJ, USA: Englewood Cliffs, N.J., Prentice-Hall, 1966.
- Justin Clarke-Doane and Dan Baras. Modal security. *Philosophy and Phenomenological Research*, 102(1):162–183, 2021. doi: 10.1111/phpr.12643.
- James Van Cleve. Epistemic supervenience and the circle of belief. *The Monist*, 68(1): 90–104, 1985. doi: monist198568115.
- C. A. J. Coady. *Testimony: A Philosophical Study*. Oxford, England: Oxford University Press, 1992.
- E. J. Coffman. Misleading dispositions and the value of knowledge. *Journal of Philosophical Research*, 35:241–258, 2010. doi: 10.5840/jpr\emph{2010}13.
- Laurence Jonathan Cohen. *The Probable and the Provable*. Oxford and New York: Clarendon Press, 1977.
- David Colaço, Wesley Buckwalter, Stephen Stich, and Edouard Machery. Epistemic intuitions in fake-barn thought experiments. *Episteme*, 11(2):199–212, 2014. doi: 10.1017/epi.2014.7.
- Mark Colyvan, Helen M. Regan, and Scott Ferson. Is it a crime to belong to a reference class. *Journal of Political Philosophy*, 9(2):168–181, 2001. doi: 10.1111/1467-9760.00123.
- E. Conee and R. Feldman. The generality problem for reliabilism. *Philosophical Studies*, 89(1):1–29, 1998. doi: 10.1023/A:1004243308503.
- Earl Conee and Richard Feldman. *Evidentialism: Essays in Epistemology*. Oxford, England: Oxford University Press, 2004.

- Fabrice Correia and Benjamin Schnieder. *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press, 2012.
- Edward Craig. *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford University Press, 1990.
- Jonathan Dancy. *An Introduction to Contemporary Epistemology*. Blackwell, 1985.
- Keith DeRose. Solving the skeptical problem. *Philosophical Review*, 104(1):1–52, 1995. doi: 10.2307/2186011.
- Keith DeRose. Assertion, knowledge, and context. *Philosophical Review*, 111(2):167–203, 2002. doi: 10.1215/00318108-111-2-167.
- Julien Dutant. The legend of the justified true belief analysis. *Philosophical Perspectives*, 29(1):95–145, 2015. doi: 10.1111/phpe.12061.
- Philip A. Ebert, Martin Smith, and Ian Durbach. Lottery judgments: A philosophical and experimental study. *Philosophical Psychology*, 31(1):110–138, 2018. doi: 10.1080/09515089.2017.1367767.
- Miguel Egler. Why understanding-why is contrastive. *Synthese*, 199(3-4):6061–6083, 2021.
- Samuel Z. Elgin. Merely partial definition and the analysis of knowledge. *Synthese*, 198 (Suppl 7):1481–1505, 2018. doi: 10.1007/s11229-018-1846-0.
- Pascal Engel. In what sense is knowledge the norm of assertion? *Grazer Philosophische Studien*, 77(1):45–59, 2008. doi: 10.1163/18756735-90000843.
- David Enoch and Talia Fisher. Sense and "sensitivity": Epistemic and instrumental approaches to statistical evidence. *Stanford law review*, 67(3):557–611, 2015.
- David Enoch, Levi Spectre, and Talia Fisher. Statistical evidence, sensitivity, and the legal value of knowledge. *Philosophy and Public Affairs*, 40(3):197–224, 2012. doi: 10.1111/papa.12000.
- David Enoch, Talia Fisher, and Levi Spectre. Does legal epistemology rest on a mistake? on fetishism, two-tier system design, and conscientious fact-finding. *Philosophical Issues*, 31(1):85–103, 2021. doi: 10.1111/phis.12203.
- David Faraci. Groundwork for an explanationist account of epistemic coincidence. *Philosophers' Imprint*, 19, 2019.
- Kit Fine. Essence and modality. *Philosophical Perspectives*, 8(Logic and Language): 1–16, 1994. doi: 10.2307/2214160.
- Jerry A. Fodor. *Concepts: Where Cognitive Science Went Wrong*. Oxford University Press, 1998.

- Richard Foley. *The Theory of Epistemic Rationality*. Harvard University Press, 1987.
- Harry Frankfurt. Alternate possibilities and moral responsibility. *Journal of Philosophy*, 66(23):829, 1969. doi: 10.2307/2023833.
- Giada Fratantonio. Evidence, risk, and proof paradoxes: Pessimism about the epistemic project. *International Journal of Evidence and Proof*, 2021.
- Craig French. Does propositional seeing entail propositional knowledge? *Theoria*, 78(2): 115–127, 2012. doi: 10.1111/j.1755-2567.2012.01130.x.
- Elizabeth Fricker. Is knowing a state of mind? the case against. In Duncan Pritchard and Patrick Greenough, editors, *Williamson on Knowledge*. Oxford: Oxford University Press, 2009.
- Georgi Gardiner. Legal epistemology. In *Oxford Bibliographies Online*. 2019.
- Georgi Gardiner. Profiling and proof: Are statistics safe? *Philosophy (London)*, 95(2): 161–183, 2020.
- Georgi Gardiner. Legal burdens of proof and statistical evidence. In James Chase and David Coady, editors, *The Routledge Handbook of Applied Epistemology*. Routledge, forthcoming.
- Mikkel Gerken. *On Folk Epistemology. How We Think and Talk About Knowledge*. Oxford: Oxford University Press, 2017.
- Edmund L. Gettier. Is justified true belief knowledge? *Analysis*, 23(6):121–123, 1963. doi: 10.1093/analys/23.6.121.
- Sanford Goldberg. Epistemic entitlement and luck. *Philosophy and Phenomenological Research*, 91(2):273–302, 2015. doi: 10.1111/phpr.12083.
- Sanford Goldberg. Testimonial knowledge through unsafe testimony. *Analysis*, 65(4): 302–311, 2005. doi: 10.1093/analys/65.4.302.
- Sanford Goldberg. *Relying on Others: An Essay in Epistemology*. Oxford University Press, 2010.
- Alan H. Goldman. An explanatory analysis of knowledge. *American Philosophical Quarterly*, 21(1):101–108, 1984.
- Alvin Goldman. What is justified belief? In George Pappas, editor, *Justification and Knowledge*, pages 1–25. Boston: D. Reidel, 1979.
- Alvin Goldman. Immediate justification and process reliabilism. In Quentin Smith, editor, *Epistemology: New Essays*, pages 63–82. Oxford University Press, 2008.

- Alvin Goldman and Bob Beddor. Reliabilist epistemology. In *Stanford Encyclopedia of Philosophy*. 2021.
- Alvin Goldman and Cailin O'Connor. Social Epistemology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.
- Alvin I. Goldman. A causal theory of knowing. *Journal of Philosophy*, 64(12):357–372, 1967. doi: 10.2307/2024268.
- Alvin I. Goldman. Discrimination and perceptual knowledge. *Journal of Philosophy*, 73 (November):771–791, 1976. doi: 10.2307/2025679.
- Alvin Ira Goldman. *Epistemology and Cognition*. Cambridge, MA, USA: Harvard University Press, 1986.
- Jeremy Goodman and Bernhard Salow. Taking a chance on kk. *Philosophical Studies*, 175(1):183–196, 2018. doi: 10.1007/s11098-017-0861-1.
- Peter J. Graham. Intelligent design and selective history: Two sources of purpose and plan. In Jonathan L. Kvanvig, editor, *Oxford Studies in Philosophy of Religion Volume 3*, pages 67–88. Oxford University Press, 2011.
- Peter J. Graham. Epistemic entitlement. *Noûs*, 46(3):449–482, 2012. doi: 10.1111/j.1468-0068.2010.00815.x.
- Peter J. Graham. Against transglobal reliabilism. *Philosophical Studies*, 169(3):525–535, 2014a. doi: 10.1007/s11098-013-0241-4.
- Peter J. Graham. Functions, warrant, history. In Abrol Fairweather and Owen Flanagan, editors, *Naturalizing Epistemic Virtue*, pages 15–35. Cambridge University Press, 2014b.
- John Greco. Worries about pritchard's safety. *Synthese*, 158(3):299–302, 2007. doi: 10.1007/s11229-006-9040-1.
- John Greco. *Achieving Knowledge: A Virtue-Theoretic Account of Epistemic Normativity*. Cambridge University Press, 2010.
- John Greco. A (different) virtue epistemology. *Philosophy and Phenomenological Research*, 85(1):1–26, 2012a. doi: ppr201285140.
- John Greco. *Better safe than sensitive*, page 193–206. Cambridge University Press, 2012b. doi: 10.1017/CBO9780511783630.015.
- John Greco. Safety in sosa. *Synthese*, 197(12):5147–5157, 2020a. doi: 10.1007/s11229-018-1863-z.

- John Greco. *Knowledge-Producing Abilities*, page 124–146. Cambridge University Press, 2020b. doi: 10.1017/9781108666404.006.
- John Greco and John Turri. Virtue epistemology. In *Stanford Encyclopedia of Philosophy*. 2017.
- Thomas Grundmann. Saving safety from counterexamples. *Synthese*, 197(12):5161–5185, 2018. doi: 10.1007/s11229-018-1677-z.
- Susan Haack. *Evidence Matters: Science, Proof, and Truth in the Law*. Cambridge University Press, 2014.
- Gilbert H. Harman. The inference to the best explanation. *Philosophical Review*, 74(1): 88–95, 1965. doi: 10.2307/2183532.
- John Hawthorne. The case for closure. In Matthias Steup and Ernest Sosa, editors, *Contemporary Debates in Epistemology*, pages 26–43. Blackwell, 2005.
- Allan Hazlett. The myth of factive verbs. *Philosophy and Phenomenological Research*, 80(3):497–522, 2010. doi: ppr201080317.
- Brian Hedden and Mark Colyvan. Legal probabilism: A qualified defence. *Journal of Political Philosophy*, 27(4):448–468, 2019. doi: 10.1111/jopp.12180.
- Mark Heller. The simple solution to the problem of generality. *Noûs*, 29(4):501–515, 1995. ISSN 00294624, 14680068. URL <http://www.jstor.org/stable/2216284>.
- Stephen Hetherington. Knowledge can be lucky. In Matthias Steup and John Turri, editors, *Contemporary Debates in Epistemology*, page 164. Blackwell, 2013.
- Avram Hiller and Ram Neta. Safety and epistemic luck. *Synthese*, 158(3):303–313, 2007. doi: 10.1007/s11229-006-9041-0.
- Jaakko Hirvelä. Global safety: How to deal with necessary truths. *Synthese*, 196(3): 1167–1186, 2019. doi: 10.1007/s11229-017-1511-z.
- Christopher Hitchcock. Probabilistic Causation. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition, 2021.
- H. L. Ho. *A Philosophy of Evidence Law: Justice in the Search for Truth*. Oxford University Press, 2008.
- Hock Lai Ho. The Legal Concept of Evidence. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.

- Wesley H. Holliday. Epistemic closure and epistemic logic i: Relevant alternatives and subjunctivism. *Journal of Philosophical Logic*, 44(1):1–62, 2015. doi: 10.1007/s10992-014-9338-2.
- Terence E. Horgan. From supervenience to superdupervenience: Meeting the demands of a material world. *Mind*, 102(408):555–86, 1993. doi: 10.1093/mind/102.408.555.
- Joachim Horvath and Jennifer Nado. Knowledge and normality. *Synthese*, 198(12): 11673–11694, 2021. doi: 10.1007/s11229-020-02823-9.
- Jonathan Ichikawa and C. S. I. Jenkins. On putting knowledge 'first'. In Joseph Adam Carter, Emma C. Gordon, and Benjamin Jarvis, editors, *Knowledge First: Approaches in Epistemology and Mind*. Oxford University Press, 2017.
- Jonathan Ichikawa and Matthias Steup. The analysis of knowledge. In *Stanford Encyclopedia of Philosophy*. 2014.
- Jonathan Jenkins Ichikawa. Justification is potential knowledge. *Canadian Journal of Philosophy*, 44(2):184–206, 2014. doi: 10.1080/00455091.2014.923240.
- Jonathan Jenkins Ichikawa and Matthias Steup. The Analysis of Knowledge. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2018 edition, 2018.
- Elizabeth G. Jackson. The relationship between belief and credence. *Philosophy Compass*, 15(6):1–13, 2020. doi: 10.1111/phc3.12668.
- Frank Jackson. *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford University Press, 1998.
- C. S. Jenkins. Knowledge and explanation. *Canadian Journal of Philosophy*, 36(2):137–164, 2006. doi: cjphil20063625.
- Daniel Kahneman, Paul Slovic, and Amos Tversky. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, 1982.
- Thomas Kelly. Evidence. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2016 edition, 2016.
- Christoph Kelp. Knowledge and safety. *Journal of Philosophical Research*, 34:21–31, 2009. doi: 10.5840/jpr\emph{2009}1.
- Christoph Kelp. In defence of virtue epistemology. *Synthese*, 179(3):409–433, 2011. doi: 10.1007/s11229-009-9681-y.
- Christoph Kelp. Knowledge: The safe-apt view. *Australasian Journal of Philosophy*, 91(2):265–278, 2013. doi: 10.1080/00048402.2012.673726.

- Christoph Kelp. Epistemic frankfurt cases revisited. *American Philosophical Quarterly*, 53(1):27–37, 2016.
- Christoph Kelp. Knowledge first virtue epistemology. In Adam Carter, Emma Gordon, and Benjamin Jarvis, editors, *Knowledge First: Approaches in Epistemology and Mind*. Oxford University Press, 2017.
- Christoph Kelp. *Good Thinking: A Knowledge First Virtue Epistemology*. London, UK: Routledge, 2018.
- Christoph Kelp and John Greco. *Virtue Theoretic Epistemology: New Methods and Approaches*. Cambridge University Press, 2020.
- Jaegwon Kim. *Supervenience and Mind: Selected Philosophical Essays*. Cambridge University Press, 1993.
- Daniel Z. Korman and Dustin Locke. Against minimalist responses to moral debunking arguments. *Oxford Studies in Metaethics*, 15:309–332, 2020.
- Daniel Z. Korman and Dustin Locke. An explanationist account of genealogical defeat. *Philosophy and Phenomenological Research*, forthcoming. doi: 10.1111/phpr.12848.
- Hilary Kornblith. *Knowledge and its Place in Nature*. Oxford University Press, 2002.
- Saul A. Kripke. Nozick on knowledge. In Saul A. Kripke, editor, *Philosophical Troubles. Collected Papers Vol I*. Oxford University Press, 2011.
- Jonathan L. Kvanvig. *The Value of Knowledge and the Pursuit of Understanding*. Cambridge University Press, 2003.
- Jennifer Lackey. Epistemic luck [review]. *Philosophical Quarterly*, 56(223):284–289, 2006. doi: 10.1111/j.1467-9213.2006.00443.x.
- Jennifer Lackey. Why we don't deserve credit for everything we know. *Synthese*, 158(3): 345–361, 2007. doi: 10.1007/s11229-006-9044-x.
- Jennifer Lackey. Knowledge and credit. *Philosophical Studies*, 142(1):27–42, 2009. doi: 10.1007/s11098-008-9304-3.
- Jennifer Lackey. Norms of criminal conviction. *Wiley: Philosophical Issues*, 31(1):188–209, 2021. doi: 10.1111/phis.12199.
- Maria Lasonen-Aarnio. Unreasonable knowledge. *Philosophical Perspectives*, 24(1): 1–21, 2010. doi: 10.1111/j.1520-8583.2010.00183.x.
- Keith Lehrer. *Theory of Knowledge*. Westview Press, 1990.
- David Lewis. *On the Plurality of Worlds*. Wiley-Blackwell, 1986.

- David Kellogg Lewis. *Counterfactuals*. Cambridge, MA, USA: Blackwell, 1973.
- Peter Lipton. *Inference to the Best Explanation*. London and New York: Routledge/Taylor and Francis Group, 1991.
- Clayton Littlejohn. Fake barns and false dilemmas. *Episteme*, 11(4):369–389, 2014. doi: 10.1017/epi.2014.24.
- Clayton Littlejohn. How and why knowledge is first. In A. Carter, E. Gordon, and B. Jarvis, editors, *Knowledge First*, pages 19–45. Oxford University Press, 2017.
- Clayton Littlejohn. Truth, knowledge, and the standard of proof in criminal law. *Synthese*, 197(12):5253–5286, forthcominga. doi: 10.1007/s11229-017-1608-4.
- Clayton Littlejohn. A plea for epistemic excuses. In Fabian Dorsch Julien Dutant, editor, *The New Evil Demon Problem*. Oxford University Press, forthcomingb.
- Clayton Littlejohn and J. Adam Carter. *This is Epistemology: An Introduction*. Wiley-Blackwell, 2021.
- Clayton Littlejohn and Julien Dutant. Justification, knowledge, and normality. *Philosophical Studies*, 177(6):1593–1609, 2020. doi: 10.1007/s11098-019-01276-2.
- Elizabeth F. Loftus. *Eyewitness testimony*. Harvard University Press, Cambridge, Massachusetts, 1996. ISBN 0674287770;9780674287778;.
- Elizabeth F. Loftus. Eyewitness testimony. *Applied cognitive psychology*, 33(4):498–503, 2019.
- William G. Lycan. On the gettier problem problem. In Stephen Hetherington, editor, *Epistemology Futures*, pages 148–168. Oxford University Press, 2006.
- Jack Lyons. Methods, processes and knowledge. In Luis Olivera, editor, *Externalism in Epistemology*. 2021.
- Jack C. Lyons. *Perception and Basic Beliefs: Zombies, Modules and the Problem of the External World*. Oxford University Press, 2009.
- Jack C. Lyons. Algorithm and parameters: Solving the generality problem for reliabilism. *Philosophical Review*, 128(4):463–509, 2019. doi: 10.1215/00318108-7697876.
- Anne Ruth Mackor and Peter Van Koppen. Scenario theory about evidence in criminal law. In Giovanni Tuzet Christian Dahlman, Alex Stein, editor, *Philosophical Foundations of Evidence Law*, pages 215–230. Oxford, UK: Oxford University Press, 2021.
- Anna-Sofia Maurin. Grounding and metaphysical explanation: It's complicated. *Philosophical Studies*, 176(6):1573–1594, 2019. doi: 10.1007/s11098-018-1080-0.

- Aidan McGlynn. The problem of true-true counterfactuals. *Analysis*, 72(2):276–285, 2012. doi: 10.1093/analys/ans046.
- Aidan McGlynn. *Knowledge First?* Palgrave Macmillian, 2014.
- Brian McLaughlin and Karen Bennett. Supervenience. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2021 edition, 2021.
- Tristram McPherson. Supervenience in Ethics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Summer 2022 edition, 2022.
- Alan Millar. Reasons for belief, perception, and reflective knowledge. *Aristotelian Society Supplementary Volume*, 88(1):1–19, 2014. doi: 10.1111/supa.2014.88.issue-1.
- Alexander Miller. *Contemporary Metaethics: An Introduction*. Polity, 2013.
- Ruth Garrett Millikan. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. MIT Press, 1984.
- Ruth Garrett Millikan. *White Queen Psychology and Other Essays for Alice*. MIT Press, 1993.
- Lisa Miracchi. Competence to know. *Philosophical Studies*, 172(1):29–56, 2015. doi: 10.1007/s11098-014-0325-9.
- Sarah Moss. Pragmatic encroachment and legal proof. *Wiley: Philosophical Issues*, 31(1):258–279, 2021. doi: 10.1111/phs.12207.
- Sarah Moss. Knowledge and legal proof. *Oxford Studies in Epistemology*, forthcoming.
- Juan Comesa na. Unsafe knowledge. *Synthese*, 146(3):395–404, 2005. doi: 10.1007/s11229-004-6213-7.
- Juan Comesa na. Knowledge and subjunctive conditionals. *Philosophy Compass*, 2(6):781–791, 2007. doi: 10.1111/j.1747-9991.2007.00076.x.
- Jennifer Nagel. Knowledge as a mental state. *Oxford Studies in Epistemology*, 4:275–310, 2013.
- Charles Nesson. Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, pages 1187–1225, 1979a.
- Charles R. Nesson. Reasonable doubt and permissive inferences: The value of complexity. *Harvard law review*, 92(6):1187–1225, 1979b.
- Ram Neta. S knows that p. *Noûs*, 36(4):663–681, 2002. doi: 10.1111/1468-0068.00406.

- Ram Neta and Guy Rohrbaugh. Luminosity and the safety of knowledge. *Pacific Philosophical Quarterly*, 85(4):396–406, 2004. doi: 10.1111/j.1468-0114.2004.00207.x.
- Jon Newman. Beyond reasonable doubt. *N.Y.U. Law Review*, 68(5):979–1002, 1993.
- Daniel Nolan. Hyperintensional metaphysics. *Philosophical Studies*, 171(1):149–160, 2014. doi: 10.1007/s11098-013-0251-2.
- Robert Nozick. *Philosophical Explanations*. Harvard University Press, 1981.
- Peter Pagin and Neri Marsili. Assertion. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.
- Michael S. Pardo. The gettier problem and legal proof: Michael s. pardo. *Legal Theory*, 16(1):37–57, 2010. doi: 10.1017/S1352325210000054.
- Michael S. Pardo. Safety vs. sensitivity: Possible worlds and the law of evidence. *Legal Theory*, 24(1):50–75, 2018. doi: 10.1017/s1352325218000010.
- Michael S. Pardo. The paradoxes of legal proof: A critical guide. *Boston University law review*, 99(1):233–290, 2019.
- Michael S. Pardo and Ronald J. Allen. Juridical proof and the best explanation. *Law and Philosophy*, 27(3):223–268, 2008. ISSN 01675249, 15730522. URL <http://www.jstor.org/stable/27652649>.
- Andrew Peet and Eli Pitcovski. Normal knowledge: Toward an explanation-based theory of knowledge. *Journal of Philosophy*, 115(3):141–157, 2018. doi: 10.5840/jphil201811539.
- N. Pennington and R. Hastie. Cognitive and social processes in decision making. In Lauren Resnick, Levine B., M. John, Stephanie Teasley, and D., editors, *Perspectives on Socially Shared Cognition*. American Psychological Association, 1991.
- Nancy Pennington and Reid Hastie. Reasoning in explanation-based decision making. *Cognition*, 49(1):123–163, 1993.
- Federico Picinali. The presumption of innocence: A deflationary account. *Modern law review*, 84(4):708–739, 2021.
- Alvin Plantinga. Warrant and designing agents: A reply to james taylor. *Philosophical Studies*, 64(2):203–215, 1991. doi: 10.1007/BF00404829.
- Alvin Plantinga. *Warrant and Proper Function*. Oxford University Press, 1993.
- Richard Posner. *The Economic Analysis of Law*. Brown and Company, 1973.
- Duncan Pritchard. *Epistemic Luck*. Oxford University Press UK, 2005.

- Duncan Pritchard. Anti-luck epistemology. *Synthese*, 158(3):277–297, 2007. doi: 10.1007/s11229-006-9039-7.
- Duncan Pritchard. Sensitivity, safety, and anti-luck epistemology. In John Greco, editor, *The Oxford Handbook of Skepticism*. Oxford University Press, 2008.
- Duncan Pritchard. Safety-based epistemology: Whither now? *Journal of Philosophical Research*, 34:33–45, 2009. doi: jpr\emph{2009}2.
- Duncan Pritchard. *The Nature and Value of Knowledge: Three Investigations: Three Investigations*. Oxford University Press, 2010.
- Duncan Pritchard. Anti-luck virtue epistemology. *Journal of Philosophy*, 109(3):247–279, 2012a. doi: 10.5840/jphil201210939.
- Duncan Pritchard. *Epistemological Disjunctivism*. Oxford University Press, 2012b.
- Duncan Pritchard. Knowledge cannot be lucky. In Matthias Steup and John Turri, editors, *Contemporary Debates in Epistemology*. Blackwell, 2013.
- Duncan Pritchard. Anti-luck epistemology and the gettier problem. *Philosophical Studies*, 172(1):93–111, 2015. doi: 10.1007/s11098-014-0374-0.
- Duncan Pritchard. Epistemic risk. *Journal of Philosophy*, 113(11):550–571, 2016. doi: 10.5840/jphil20161131137.
- Duncan Pritchard. Anti-luck virtue epistemology and epistemic defeat. *Synthese*, 195(7): 3065–3077, 2018a. doi: 10.1007/s11229-016-1074-4.
- Duncan Pritchard. Legal risk, legal evidence and the arithmetic of criminal justice. *Jurisprudence*, 9(1):108–119, 2018b. doi: 10.1080/20403313.2017.1352323.
- Duncan Pritchard. *Anti-risk Virtue Epistemology*, page 203–224. Cambridge University Press, 2020. doi: 10.1017/9781108666404.010.
- Duncan Pritchard. Varieties of epistemic risk. *Acta Analytica*, 37(1):9–23, 2022. doi: 10.1007/s12136-021-00489-7.
- Duncan Pritchard and Patrick Greenough. *Williamson on Knowledge*. Oxford: Oxford University Press, 2009.
- Duncan Pritchard and Jesper Kallestrup. Epistemic supervenience, anti-individualism and knowledge first epistemology. In Joseph Adam Carter, Emma C. Gordon, and Benjamin Jarvis, editors, *Knowledge First: Approaches in Epistemology and Mind*. Oxford University Press, 2017.
- James Pryor. Comments on sosa’s relevant alternatives, contextualism included. *Philosophical Studies*, 119(1-2):67–72, 2004. doi: 10.1023/B:PHIL.0000029350.30567.bb.

- Amit Pundik. The epistemology of statistical evidence. *The international journal of evidence & proof*, 15(2):117–143, 2011.
- Christian Dahlman & Amit Pundik. Problem with naked statistical evidence. In Giovanni Tuzet Christian Dahlman, Alex Stein, editor, *Philosophical Foundations of Evidence Law*, pages 333–344. Oxford, UK: Oxford University Press, 2021.
- Dani Rabinowitz. The safety condition for knowledge. In Ted Poston, editor, *Internet Encyclopedia of Philosophy*. 2011.
- Colin Radford. Knowledge: By examples. *Analysis*, 27(1):1, 1966. doi: 10.2307/3326979.
- Chris Ranalli. Luck, propositional perception, and the entailment thesis. *Synthese*, 191(6):1223–1247, 2014. doi: 10.1007/s11229-013-0323-z.
- Mike Redmayne. Exploring the proof paradoxes. *Legal Theory*, 14(4):281–309, 2008. doi: 10.1017/s1352325208080117.
- Steven L. Reynolds. Justification as the appearance of knowledge. *Philosophical Studies*, 163(2):367–383, 2013. doi: 10.1007/s11098-011-9820-4.
- Steven Rieber. Skepticism and contrastive explanation. *Noûs*, 32(2):189–204, 1998. doi: 10.1111/0029-4624.00096.
- Wayne Riggs. Two problems of easy credit. *Synthese*, 169(1):201–216, 2009. doi: 10.1007/s11229-008-9342-6.
- Jeffrey Roland and Jon Cogburn. Anti-luck epistemologies and necessary truths. *Philosophia*, 39(3):547–561, 2011. doi: 10.1007/s11406-010-9295-0.
- Jasmine B. Gonzales Rose. Race, evidence and epistemic injustice. In Giovanni Tuzet Christian Dahlman, Alex Stein, editor, *Philosophical Foundations of Evidence Law*, pages 380–390. Oxford, UK: Oxford University Press, 2021.
- Gideon Rosen. Metaphysical dependence: Grounding and reduction. In Bob Hale and Aviv Hoffmann, editors, *Modality: Metaphysics, Logic, and Epistemology*, pages 109–135. Oxford University Press, 2010.
- Lewis Ross. Justice in epistemic gaps: The ‘proof paradox’ revisited. *Philosophical Issues*, 31(1):315–333, 2021a. doi: 10.1111/phis.12193.
- Lewis Ross. Rehabilitating statistical evidence. *Philosophy and Phenomenological Research*, 102(1):3–23, 2021b.
- Lewis D. Ross. Legal proof and statistical conjunctions. *Philosophical Studies*, 178(6): 2021–2041, 2020a. doi: 10.1007/s11098-020-01521-z.

- Lewis D. Ross. Recent work on the proof paradox. *Philosophy Compass*, 15(6), 2020b. doi: 10.1111/phc3.12667.
- Andrea Roth. Safety in numbers? deciding when dna alone is enough to convict. *New York University law review (1950)*, 85(4):1130, 2010.
- Sherrilyn Roush. *Tracking Truth: Knowledge, Evidence, and Science*. Oxford University Press, 2005.
- R. M. Sainsbury. Easy possibilities. *Philosophy and Phenomenological Research*, 57(4): 907–919, 1997. doi: 10.2307/2953809.
- Jonathan Schaffer. Knowledge in the image of assertion. *Philosophical Issues*, 18(1): 1–19, 2008. doi: 10.1111/j.1533-6077.2008.00134.x.
- Susanna Schellenberg. *The Unity of Perception: Content, Consciousness, Evidence*. Oxford: Oxford University Press, 2018.
- Benjamin Schnieder. A logic for 'because'. *Review of Symbolic Logic*, 4(3):445–465, 2011. doi: 10.1017/s1755020311000104.
- Yafeng Shan. Contrastivism and non-contrastivism in scientific explanation. *Philosophy Compass*, 14(8):e12613, 2019. doi: 10.1111/phc3.12613.
- Robert K. Shope. *The Analysis of Knowing: A Decade of Research*. Princeton University Press, 1983.
- Theodore Sider. *Writing the Book of the World*. Oxford, England: Oxford University Press, 2011.
- Susanna Siegel. The Contents of Perception. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2021 edition, 2021.
- Mona Simion. Assertion: Knowledge is enough. *Synthese*, 193(10), 2016a. doi: 10.1007/s11229-015-0914-y.
- Mona Simion. Perception, history and benefit. *Episteme*, 13(1):61–76, 2016b. doi: 10.1017/epi.2015.56.
- Mona Simion. Knowledge-first functionalism. *Philosophical Issues*, 29(1):254–267, 2019. doi: 10.1111/phis.12152.
- Mona Simion and Christoph Kelp. Assertion: The constitutive norms view. In Sanford Goldberg, editor, *The Oxford Handbook of Assertion*. Oxford University Press, 2019.
- Martin Smith. What else justification could be. *Noûs*, 44(1):10–31, 2010. doi: 10.1111/j.1468-0068.2009.00729.x.

- Martin Smith. *Between Probability and Certainty: What Justifies Belief*. Oxford University Press UK, 2016.
- Martin Smith. The cost of treating knowledge as a mental state. In A. Carter, E. Gordon, and B. Jarvis, editors, *Knowledge First, Approaches to Epistemology and Mind*, pages 95–112. Oxford University Press, 2017.
- Martin Smith. When does evidence suffice for conviction? *Mind*, 127(508):1193–1218, 2018. doi: 10.1093/mind/fzx026.
- Ernest Sosa. *Knowledge in Perspective: Selected Essays in Epistemology*. Cambridge: Cambridge University Press, 1991.
- Ernest Sosa. Review: Proper functionalism and virtue epistemology. *Noûs*, 27(1):51–65, 1993. doi: 10.2307/2215895.
- Ernest Sosa. How to defeat opposition to moore. *Philosophical Perspectives*, 13:137–49, 1999. doi: 10.1111/0029-4624.33.s13.7.
- Ernest Sosa. Response to plantinga. In John Greco, editor, *Ernest Sosa: And His Critics*. Oxford: Blackwell, 2004.
- Ernest Sosa. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford University Press, 2007.
- Ernest Sosa. How competence matters in epistemology. *Philosophical Perspectives*, 24(1):465–475, 2010. doi: 10.1111/j.1520-8583.2010.00200.x.
- Ernest Sosa. *Judgement and Agency*. Oxford: Oxford University Press, 2015.
- Ernest Sosa. *Epistemology*. Princeton: Princeton University Press, 2017.
- Ernest Sosa. *Epistemic Explanations*. Oxford: Oxford University Press, 2021.
- Mark Spottswood. Paradoxes of proof. In Giovanni Tuzet Christian Dahlman, Alex Stein, editor, *Philosophical Foundations of Evidence Law*, pages 317–331. Oxford, UK: Oxford University Press, 2021.
- Jason Stanley. *Knowledge and Practical Interests*. Oxford University Press, 2005.
- Daniel Stoljar. Physicalism and phenomenal concepts. *Mind and Language*, 20(2):296–302, 2005. doi: 10.1111/j.0268-1064.2005.00296.x.
- Michael Strevens. *Depth: An Account of Scientific Explanation*. Harvard University Press, 2008.
- Jonathan Sutton. *Without Justification*. MIT Press, 2007.

- Kurt Sylvan. Knowledge as a non-normative relation. *Philosophy and Phenomenological Research*, 97(1):190–222, 2018. doi: 10.1111/phpr.12371.
- Judith Jarvis Thomson. Liability and individualized evidence. *Law and Contemporary Problems*, 49(3):199–219, 1986. ISSN 00239186. URL <http://www.jstor.org/stable/1191633>.
- Laurence H. Tribe. Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, 84(6):1329–1393, 1971. ISSN 0017811X. URL <http://www.jstor.org/stable/1339610>.
- John Turri. Epistemic supervenience. In Jonathan Dancy, Ernest Sosa, and Matthias Steup, editors, *Blackwell Companion to Epistemology*, 2 ed. Blackwell, 2010.
- John Turri. A new paradigm for epistemology from reliabilism to abilism. *Ergo: An Open Access Journal of Philosophy*, 3, 2016. doi: 10.3998/ergo.12405314.0003.008.
- Peter Unger. An analysis of factual knowledge. *Journal of Philosophy*, 65(6):157–170, 1968. doi: 10.2307/2024203.
- Rafal Urbaniak and Marcello Di Bello. Legal Probabilism. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2021 edition, 2021.
- Barbara Vetter. Dispositions without conditionals. *Mind*, 123(489):129–156, 2014. doi: 10.1093/mind/fzu032.
- Barbara Vetter. *Potentiality: From Dispositions to Modality*. Oxford, England and New York, NY, USA: Oxford University Press, 2015.
- Brian Weatherson. What good are counterexamples? *Philosophical Studies*, 115(1):1–31, 2003. doi: 10.1023/A:1024961917413.
- Ralph Wedgwood. The internalist virtue theory of knowledge. *Synthese*, 197(12):5357–5378, 2020. doi: 10.1007/s11229-018-1707-x.
- Dennis Whitcomb. Factivity without safety. *Pacific Philosophical Quarterly*, 89(1):143–149, 2008. doi: 10.1111/j.1468-0114.2008.00313.x.
- Timothy Williamson. *Vagueness*. London and New York: Routledge, 1994.
- Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, 2000.
- Timothy Williamson. Reply to cassam. In Duncan Pritchard and Patrick Greenough, editors, *Williamson on Knowledge*. Oxford University Press, 2009a.
- Timothy Williamson. Probability and danger. In *Amherst Lecture in Philosophy*, pages 1–35. 2009b.

- Timothy Williamson. Justifications, excuses, and sceptical scenarios. In Fabian Dorsch and Julien Dutant, editors, *The New Evil Demon*. Oxford: Oxford University Press, forthcoming.
- Jessica M. Wilson. No work for a theory of grounding. *Inquiry: An Interdisciplinary Journal of Philosophy*, 57(5-6):535–579, 2014. doi: 10.1080/0020174x.2014.907542.
- Sara Worley. Determination and mental causation. *Erkenntnis*, 46(3):281–304, 1997. doi: 10.1023/A:1005301816477.
- Larry Wright. Functions. *Philosophical Review*, 82(2):139–168, 1973. doi: 10.2307/2183766.
- Linda Zagzebski. The inescapability of gettier problems. *Philosophical Quarterly*, 44(174):65–73, 1994. doi: 10.2307/2220147.
- Nick Zangwill. Epistemic/non-epistemic dependence. *Noûs*, pages 836–857, 2018. doi: 10.1111/nous.12197.