

# The Fragile Estate:

Essays on luminosity, normativity  
and metaphilosophy

Amia Srinivasan

All Souls College, Oxford

Faculty of Philosophy

A thesis submitted for the degree of DPhil in Philosophy

Trinity Term, 2013

# The Fragile Estate:

Essays on luminosity, normativity and metaphilosophy

Amia Srinivasan

All Souls College, Oxford; Faculty of Philosophy

A thesis submitted for the degree of DPhil in Philosophy: Trinity Term, 2013

## ABSTRACT

This dissertation is a set of three essays connected by the common theme of our *epistemic fragility*: the way in which our knowledge – of our own minds, of whether we are in violation of the epistemic and ethical norms, and of the philosophical truths themselves – is hostage to forces outside our control. The first essay, “Are We Luminous?”, is a recasting and defence of Timothy Williamson’s argument that there are no non-trivial conditions such that we are in a position to know we are in them whenever we are in them. Crucial to seeing why Williamson’s anti-luminosity argument succeeds, *pace* various critics, is recognising that the issue is largely an empirical one. It is in part because of the kind of creatures we are – specifically, creatures with coarse-grained doxastic dispositions – that nothing of interest, for us, is luminous. In the second essay, “What’s in a Norm?”, I argue that such an Anti-Cartesian view in turn demands that epistemologists and ethicists accept the ubiquity of *normative luck*, the phenomenon whereby agents fail to do what they ought because of non-culpable ignorance. Those who find such a view intolerable – many epistemic internalists and ethical subjectivists – have the option of cleaving to the Cartesian orthodoxy by endorsing an anti-realist metanormativity. The third essay, “The Archimedean Urge”, is a critical discussion of genealogical scepticism about philosophical judgment, including evolutionary debunking arguments and experimentally-motivated attacks. Although such genealogical scepticism often purports to stand outside philosophy – in the neutral terrains of science or common sense – it tacitly relies on various first-order epistemic judgments. The upshot is two-fold. First, genealogical scepticism risks self-defeat, impugning commitment to its own premises. Second, philosophers have at their disposal epistemological resources to fend off genealogical scepticism: namely, an epistemology that takes seriously the role that luck plays in the acquisition of philosophical knowledge.

Word Count: 52,285

## Acknowledgments

I have received help and guidance from many people in the writing of these essays; their names are listed at the end of each.

For his supervision and friendship over many years, I am much indebted to John Hawthorne.

I am also grateful to Tim Williamson for his wisdom, philosophical and otherwise, over the last year.

For seeing me through to the end, my love to Daniela, Kate, Mary, Alex, Hussein, Jane, Daniel and (most of all) Paul.

## Contents

Introduction: the fragile estate	I
Essay One: Are We Luminous?	13
Essay Two: What's in a Norm?	47
Essay Three: The Archimedean Urge	100

## Introduction: the fragile estate

I take the title of my dissertation from a phrase used by the classicist E.R. Dodds in his wonderful essay “On Misunderstanding the ‘Oedipus Rex’” (1966). In that essay, Dodds argues that Oedipus’ downfall is due not to any ethical flaw or failing on his part, as is so often supposed, but is instead simply a matter of bad luck. For Oedipus is not in a position to know the morally relevant facts of his situation: that the woman he marries is his mother, that the man he kills is his father. Herein, in Dodds’ view, lies the genuinely tragic nature of *Oedipus*. A downfall that results from a moral flaw is no tragedy, but the story of a great man brought down by forces outside his control is. Such a story calls our attention to, and evokes pity for, the “fragile estate of man” (ibid., 40).

My dissertation is a set of three essays connected by the common theme of our, in Dodds’ sense, *fragility*: the way in which our knowledge – of our own minds, of whether we are in violation of the epistemic and ethical norms, and of the philosophical truths themselves – is hostage to forces outside our control. The first essay, “Are We Luminous?”, is a recasting and defence of Timothy Williamson’s argument for the conclusion that there are no non-trivial conditions such that we are in a position to know we are in them whenever we are in them. In the second essay, “What’s in a Norm?”, I argue that an Anti-Cartesian view of the mind – a view suggested by the anti-luminosity argument – requires epistemologists and ethicists to accept the ubiquity of *normative luck*, the phenomenon whereby agents fail to do what they ought through non-culpable ignorance. The third and final essay, “The Archimedean Urge”, is a critical discussion of various forms of genealogically-motivated scepticism about philosophical judgment, including evolutionary debunking arguments and experimental attacks on intuition. Although such genealogical scepticism often purports to stand outside philosophy – in the neutral terrains of science or common sense – it tacitly relies on various

first-order epistemic judgments. The upshot is two-fold. First, genealogical scepticism risks self-defeat, impugning commitment to its own premises. Second, philosophers have at their disposal resources to fend off genealogical scepticism: namely, an epistemology that takes seriously the role that luck plays in the acquisition of philosophical knowledge.

### Essay One: Are We Luminous?

In *Knowledge and its Limits* (2000, ch. 4), Williamson argues that there are no non-trivial conditions, including mental state conditions, that are *luminous*, where a condition is luminous just in case whenever one is in it, one is in a position to know one is in it. Williamson does so by constructing a thought experiment involving an agent S who, over the course of a morning, we are to suppose is very slowly warming up, from feeling freezing cold at dawn to feeling very hot by noon. On the assumption that the condition of *feeling cold* is luminous, and that S is carefully attending to her feelings of coldness, it follows that S knows she is cold in every moment that she is cold. But this, together with a certain *margin-for-error* premise – according to which if S knows she is cold at one moment, then she is cold in the next – generates the conclusion that S feels cold at noon. Thus we arrive at a *reductio ad absurdum*. Williamson claims that, to avoid absurdity, we must reject the luminosity of feeling cold. But not only that. Since analogous thought experiments can be generated for any other non-trivial condition – a condition that neither always obtains nor never obtains, and that comes in degrees – we must reject luminosity altogether.

Since its appearance, the anti-luminosity argument has come under sustained attack. Defenders of the luminous overwhelmingly object to the argument's margin-for-error premise (Leitgeb 2002; Weatherson 2004; Blackson 2007; Wong 2008; Berker 2008; Ramachandran 2009; Vogel 2010; Cohen 2010; Zardini

forthcoming). Williamson himself claims that the premise follows uncontroversially from a safety condition on knowledge together with his description of the thought experiment. But luminists argue that this is not so: the margin-for-error premise either requires an implausible interpretation of the safety requirement on knowledge, or it requires other equally implausible (and soritical) assumptions.

In “Are We Luminous?” I bolster the margin-for-error premise against these attacks by recasting Williamson’s own two-part defence of it, the first part intended to work on the assumption that there is no constitutive connection between the phenomenal and the doxastic, and the second intended to work while dispensing with this assumption. *Pace* various luminists, I argue that the appeals to safety needed for Williamson’s two-part defence (the first in terms of outright belief, the second in terms of degrees of confidence) are independently plausible. I also argue that all that is needed to generate the margin-for-error premise from these safety conditions is an *empirical* assumption about the agent S in Williamson’s thought experiment: that she is, like us, a creature whose beliefs are structured by certain dispositions.

According to my favoured reconstruction, the essence of the anti-luminosity argument is this.<sup>1</sup> Suppose that S, like us, is the sort of creature for whom believing something in one situation means having, in extremely similar situations, the disposition to believe the same thing. And suppose that S is in a condition C, but in what we might call a *liminal* case of it. That is, there is an extremely similar case to the one she is in which is not a case of C. Suppose that in this liminal case S believes she is in C. Is this belief knowledge? Plausibly not. For in the very

---

<sup>1</sup> This version of the argument, in terms of outright belief, only works on the assumption that there is no constitutive connection between whether S is in C and whether S believes she is in C. A refined version of the argument in terms of degrees of confidence is required if it is to work without that assumption. For the sake of simplicity I do not go into the details of that version here; see “Are We Luminous?”, section 5 for the fully elaborated defence of the argument.

similar non-C case, S has the disposition to believe she is in C. Thus, her true belief that she is in C is rendered unreliable by a nearby untrue belief that she is in C. Thus in such a liminal case of being in C, S cannot know that she is in C.

My reconstruction thus brings out what I take to be really at stake in the dialectic about luminosity: that is, whether there are any conditions that are, *for us*, luminous. It is our empirical character – the kind of coarse-grained dispositions that structure our beliefs – that makes it plausibly the case that for us nothing of interest is luminous.

## Essay Two: What's in a Norm?

According to the Cartesian orthodoxy, the phenomenal is a realm of privileged epistemic access. But arguments from empirical psychology (Gopnik 1983; Carruthers 2011; Schwitzgebel 2008, 2011) and, as I argue in “Are We Luminous?”, armchair epistemology (Williamson 2000, ch. 4) together motivate a view on which the phenomenal is not always open to view. The rejection of the Cartesian orthodoxy has far-reaching philosophical consequences. In particular, I want to suggest, it presents normative philosophy with a serious challenge. For the putative transparency of certain phenomenal conditions has been thought by many to make them uniquely suited to feature in epistemic and ethical norms. It is often said that such norms should capture our intuitive judgments about agential culpability: whether agents are (ethically or epistemically) *blameworthy* for their actions. Norms that are tied to agents' phenomenal states – internalist epistemic norms<sup>2</sup> and subjectivist ethical norms – are often thought to have just this feature. Since agents are always in a position to know whether they are in certain phenomenal states (the Cartesian orthodoxy goes), it is impossible to blamelessly

---

<sup>2</sup> As I make clear in the essay, the kind of internalism I have in mind is what is known as ‘access’ internalism; my argument does not apply to ‘mental state’ internalism (Conee and Feldman 2001; Wedgwood 2002).

violate internalist and subjectivist norms. Thus norms that are tied to phenomenal states seem poised to vindicate an intimacy between the *normativity perspective*, from which it matters whether agents conform to norms, and the *culpability perspective*, from which it matters whether agents are blameworthy.

If we reject the Cartesian orthodoxy, however, then it turns out that even internalist and subjectivist norms cannot vindicate an intimacy between the two perspectives. For if there are no conditions such that we are always in a position to know whether we are in them, then it is possible to unknowingly, and thus blamelessly, violate any norm. This is particularly troubling news for epistemic internalists and ethical subjectivists, since Anti-Cartesianism implies that there exists no supervenience base for epistemic justification or the moral ought. But many externalists and objectivists should also be worried. For most think that some normative notion that aligns with the culpability perspective – the ‘subjective’ moral ought (as opposed to the ‘objective’ moral ought) or epistemic ‘rationality’ (as opposed to epistemic justification) – is required in a complete normative theory. But if Anti-Cartesianism is correct, then no theory can align the normative and culpability perspectives. Ethicists and epistemologists are thus faced with a dilemma. Either we cleave to Cartesianism, or we accept that the normativity and culpability perspectives can never be perfectly reconciled: normative luck is pervasive.

In “Are We Luminous?” I discuss the possibility, raised by various luminists (Weatherson 2004; Berker 2008; Ramachandran 2009), that the anti-luminosity argument doesn’t work on the assumption that there exists a constitutive connection between the phenomenal and the doxstic. For if there is a condition C such that whether S is in C is constitutively connected to S’s beliefs about whether she is in C, it might seem that whenever S is in C, she is in a position to know that she is in it. The constitutive connection between C’s obtaining and S’s beliefs

about whether C obtains seems to rule out the possibility that S could fail to be in a position to know that she is in C when she is. In “Are We Luminous?” I argue, in defence of Williamson, that this appearance is misleading. Even conditions that are constitutively connected to beliefs about whether they obtain are plausibly susceptible to the Anti-Luminosity argument, again in part because of the kind of creatures we are, empirically speaking. Nonetheless, positing such constitutive connections between the phenomenal and the doxastic is the luminist’s best hope for salvaging luminosity. It is also, I suggest in “What’s in a Norm?”, the best hope for epistemologists and ethicists wishing to deny the pervasiveness of normative luck. Such normative theorists might embrace a constitutive connection between the normative facts – what we ought or have reason to do – and our beliefs about what we ought or have reason to do. In other words, grasping the first horn of the dilemma most plausibly means endorsing a form of agent-relative constructivism about the normative realm.

Grasping the second horn of the dilemma involves accepting that the normativity and culpability perspectives can never be brought into perfect alignment. I suggest that this means embracing a more tragic conception of the normative realm, according to which luck plays an ineradicable role in our ethical and epistemic performance. It also perhaps means loosening our fixation with the normative centrality of the culpability perspective. Sometimes the most important thing is not whether we have done our best, but whether we have done well. And doing well, unlike doing one’s best, requires the cooperation of an unforgiving world.

### Essay Three: The Archimedean Urge

Recently, philosophy has been subject to sceptical attacks that take as one of their premises claims about the *genealogical contingency* of philosophical judgments. Many ethicists argue that the evolutionary origins of our moral beliefs demand

that we abandon those beliefs or adopt an anti-realist construal of their contents (Harman 1977, 1986; Singer 1981, 2005; Ruse 1985; Ruse and Wilson 1986; Gibbard 1990; Kitcher 2005, 2011; Joyce 2006, ch. 6; Street 2006, 2008, 2011; Greene 2008; Huemer 2008; Rosenberg 2011). Genealogically-motivated scepticism has also been advanced against metaphysics (Ladyman and Ross 2007), mathematical realism (Benacerraf 1973; Field 1989), logic (Cooper 2003), theism (Dennett 2006) and naturalism (Plantinga 1993, ch. 12, 1994, 2002). And the new sub-discipline of ‘experimental philosophy’ is largely devoted to arguing that philosophical judgments systematically vary with culture, gender, socioeconomic status, extent of philosophical training and so on – and thus that philosophers should not rely on their judgments in these domains (for an overview, see Knobe and Nichols 2008 and Alexander 2012).

In “The Archimedean Urge” I argue that any convincing genealogically-motivated sceptical argument against philosophical judgment must rely on various first-order epistemic premises. I show this by formulating what I take to be the five most plausible arguments for genealogical scepticism about philosophy – from (1) sensitivity, (2) explanatory inertness, (3) coincidence, (4) probability on the evidence and (5) unreliability – and highlighting each argument’s epistemic presuppositions: about explanation, evidence, justification, defeat, methods, bootstrapping and so on. This reliance on first-order epistemological premises is in spite of the claim made by many genealogical sceptics – particularly experimental philosophers – to stand outside philosophy, in the neutral terrains of science or common sense. *Pace* such critics, there is no Archimedean point from which to issue a genealogical critique of philosophy.

The upshot is two-fold. First, genealogical scepticism risks self-defeat, impugning commitment to its own premises. For it is hard to see how commitment to these epistemological premises does not itself exhibit the kind of genealogical

contingency that worries genealogical sceptics, particularly as there exists widespread and systematic variation in the acceptance of these premises by epistemologists themselves. Second, philosophers have available to them the resources required to defend against genealogical scepticism. Specifically, an epistemology that embraces the role that luck plays in the acquisition of knowledge can contribute to a metaphilosophy that is robust against genealogical attacks. For inasmuch as one takes seriously the idea that forces outside one's control can determine whether one knows, one can be sanguine about the phenomenon of *genealogical luck* – the luck of having knowledge-conducive or knowledge-hindering genealogies. As I argue in “What’s in a Norm?”, some philosophers – those inclined toward epistemic externalism and ethical objectivism – tend to embrace the role that luck plays in the normative sphere. Such philosophers have antecedent reason to embrace a metaphilosophy according to which genealogical luck plays a central role in the acquisition of philosophical knowledge. Others might be more reluctant to embrace such a metaphilosophy. Such philosophers, I suggest, might take refuge in a Pyrrhonian suspension of judgment about the ability of philosophical enquiry to produce knowledge.

I conclude the essay by discussing what I take to be the moral considerations that often motivate genealogical scepticism. There is no doubt something ethically problematic about loudly insisting that some people – philosophers, or worse, Westerners – are the beneficiaries of a knowledge-conducive genealogy, while others (non-philosophers, non-Westerners) are not. It is this moral instinct, I suspect, that pulls many toward forms of nihilism or anti-realism about philosophy: to the view that there are no philosophical truths to get right or wrong, or that the philosophical truths are in some sense parochial (e.g. just truths about our genealogically-contingent concepts). I suggest that we can respect this moral instinct while maintaining a realist metaphilosophy. First, we can distinguish between what it takes to attain philosophical knowledge, and what it

takes to be worthy of praise for attaining such knowledge; this is a particular instance of the distinction between the normativity and culpability perspectives for which I argue in “What’s in a Norm?”. Second, and more importantly, we can distinguish between epistemology and politics. For there is the question of what conditions are required to be able to know, and there is the question of how, in a just society, we should respond to other people’s claims to knowledge. The former question is the topic of this essay. But the latter, political question is also one for philosophy. In philosophy as much as elsewhere we must decide how best to govern ourselves such that we duly respect each other’s claims to knowledge. While I do not propose an answer to the question of how we might do this, I suggest that neither metaphilosophical nihilism nor anti-realism is what is called for. Instead a renewed attention to some prosaic virtues – openness, empathy, kindness – might do.

Together these essays, I hope, make a contribution to the topic of what a post-Cartesian philosophy might look like. Such a philosophy calls on us to take more seriously the fragility of our human estate and, with it, the tragic nature of the philosophical enterprise itself.

## Works Cited

- Alexander, J. (2012). *Experimental Philosophy: An Introduction*. Cambridge: Polity.
- Benacerraf, P. (1973). "Mathematical Truth". *The Journal of Philosophy*, 70(19): 661-679.
- Berker, S. (2008). "Luminosity Regained". *Philosophers' Imprint*, 8(2): 1-22.
- Blackson, T. (2007). "On Williamson's Argument for (I) in His Anti-Luminosity Argument". *Philosophy and Phenomenological Research*, 74(2): 397 – 405.
- Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Cohen, S. (2010). "Luminosity, Reliability, and the Sorites". *Philosophy and Phenomenological Research*, 81(3): 718 – 30.
- Conee, E. and Feldman, R. (2001). "Internalism Defended". *American Philosophical Quarterly*, 38(1): 1-18.
- Cooper, W.S. (2003). *The Evolution of Reason: logic as a branch of biology*. Cambridge: Cambridge University Press.
- Dennett, D. (2006). *Breaking the Spell: Religion as a Natural Phenomenon*. New York: Penguin.
- Dodds, E.R. (1966). "On Misunderstanding the 'Oedipus Rex'". *Greece & Rome*, Second Series, 13(1): 37 – 49.
- Field, H. (1989). *Realism, mathematics and modality*. Oxford: Basil Blackwell.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- Gopnik, A. (1983). "How we can know our minds: The illusion of first person knowledge of intentionality". *Brain and Behavioral Science*, 16: 1-14.
- Greene, J. (2008). "The Secret Joke of Kant's Soul", in Sinnott-Armstrong, W. (ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality, Emotions, Brain Disorders and Development*. Cambridge, MA: MIT Press, 35-80.
- Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.
- Harman, G. (1986). "Moral explanations of natural facts: can moral claims be tested against moral reality?" *Southern Journal of Philosophy*, 24: 57-68.
- Huemer, M. (2008). "Revisionary Intuitionism". *Social Philosophy and Policy*, 25: 368-392.
- Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.

- Kitcher, P. (2005). "Biology and Ethics," in Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press, ch. 6.
- Kitcher, P. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.
- Knobe, J. and Nichols, S. (eds.) (2008). *Experimental Philosophy*. Oxford: Oxford University Press.
- Ladyman, J. and Ross, D., with Spurrett, D. and Collier, J. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
- Leitgeb, H. (2002). "Review of Timothy Williamson, *Knowledge and its Limits*". *Grazer Philosophische Studien*, 65: 207-17.
- Plantinga, A. (1993). *Warrant and Proper Function*. Oxford: Oxford University Press.
- Plantinga, A. (1994). "Naturalism Defeated". Unpublished manuscript. Available at: [http://www.calvin.edu/academic/philosophy/virtual\\_library/articles/plantinga\\_alvin/naturalism\\_defeated.pdf](http://www.calvin.edu/academic/philosophy/virtual_library/articles/plantinga_alvin/naturalism_defeated.pdf)
- Plantinga, A. (2002). "The Evolutionary Argument Against Naturalism", in Beilby, J. (ed.), *Naturalism Defeated?* Ithaca: Cornell University Press, 1 – 12.
- Ramachandran, M. (2009). "Anti-Luminosity: Four Unsuccessful Strategies". *Australasian Journal of Philosophy*, 87(4): 659-673.
- Rosenberg, Alex (2011). *The Atheist's Guide to Reality: Enjoying Life Without Illusions*. New York: W. W. Norton & Co.
- Ruse, M. (1985). *Taking Darwin Seriously*. Oxford: Blackwell.
- Ruse, M. and Wilson, E.O. (1986). "Moral Philosophy as Applied Science". *Philosophy* 61: 173-92.
- Schwitzgebel, E. (2008). "The Unreliability of Naïve Introspection". *Philosophical Review*, 117: 245-273.
- Schwitzgebel, E. (2011). *Perplexities of Consciousness*. Cambridge, MA: MIT Press.
- Singer, P. (1981). *The Expanding Circle*. New York: Farrar, Straus & Giroux.
- Singer, P. (2005). "Ethics and Intuitions". *Journal of Ethics*, 9: 331-52.
- Street, S. (2006). "A Darwinian Dilemma for Realist Theories of Value". *Philosophical Studies*, 127: 109-166.
- Street, S. (2008). "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About". *Philosophical Issues*, 18: 207-228.
- Street, S. (2011). "Mind-Independence Without Mystery: Why Quasi-Realists Can't Have It Both Ways", in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics* 6. Oxford: Oxford University Press, 1-32.
- Vogel, J. (2010). "Luminosity and Indiscriminability". *Philosophical Perspectives*, 24(1): 547-572.

Weatherson, B. (2004). "Luminous Margins". *Australasian Journal of Philosophy*, 82: 373-383.

Wedgwood, R. (2002). "Internalism Explained". *Philosophy and Phenomenological Research*, 65(2): 349-369.

Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.

Wong, W. (2008). "What Williamson's Anti-Luminosity Argument Really Is". *Pacific Philosophical Quarterly*, 89: 536-543.

Zardini, E. (forthcoming). "Luminosity and determinacy". *Philosophical Studies*.

Essay One  
Are We Luminous?

## I. Introduction

In *Knowledge and Its Limits*, Timothy Williamson argues by way of a thought experiment that there are no non-trivial luminous conditions, where a condition is luminous just in case whenever one is in it, one is in a position to know one is in it (2000, ch. 4). If Williamson is right, then the common picture of the phenomenal realm as one of privileged access turns out to be a Cartesian orthodoxy from which philosophy must be cleansed. It also follows that rationality, evidence, normative obligations, and sameness of meaning – phenomena associated, for many, with privileged first-person access – are themselves non-luminous.

Given its potential to destabilise, it is little wonder that the anti-luminosity argument has come under fire since its appearance. Luminists typically attack Williamson's use of a certain *margin-for-error premise*. Williamson himself claims that the premise follows easily from a safety condition on knowledge together with his description of the thought experiment. But luminists argue that this is not so: the margin-for-error premise either requires an implausible interpretation of the safety requirement on knowledge, or it requires a plausible interpretation of the safety requirement together with other implausible, often soritical, assumptions (Leitgeb 2002; Weatherson 2004; Blackson 2007; Wong 2008; Berker 2008; Ramachandran 2009; Vogel 2010; Cohen 2010; Zardini forthcoming). Either way, the margin-for-error premise, and thus the anti-luminosity argument, is in trouble. Luminists counsel that we dismiss Williamson's argument and cleave to the luminous.

I shall argue that Williamson's controversial margin-for-error premise, *pace* the luminists, can be derived from a plausible safety condition on knowledge<sup>3</sup> together

---

<sup>3</sup> I will be not addressing those luminists who reject safety wholesale (Brueckner and Fiocco 2002; Neta and Rohrbaugh 2004; Comesaña 2005; Conee 2005), but only those who accept a safety

with a plausible empirical hypothesis about the kind of creatures we are – creatures, namely, whose beliefs are structured by certain coarse-grained dispositions. Indeed, I shall make this argument twice over. This is because some luminists have been keen to argue that the margin-for-error premise is particularly problematic on a view that maintains a constitutive connection between the phenomenal and the doxastic. So I will first argue for the margin-for-error premise from safety and empirical considerations on the assumption that no such constitutive connection obtains. I will then make a more refined argument that dispenses with that assumption: that is, an argument that applies to all (non-trivial) phenomenal conditions regardless of any constitutive connection that might obtain between them and beliefs about whether they obtain. In so doing, I hope to show the anti-luminosity argument to be robust against some of the most common criticisms. I also hope to show that a large part of what is at stake in the debate about anti-luminosity is a certain vision of what kind of creatures we are, empirically speaking.

## 2. The Anti-Luminosity Argument

Williamson aims to establish that, for almost any condition,<sup>4</sup> it is possible for a normal human to be in that condition and fail to be in a position to know that she is in it.<sup>5</sup> He attempts to do this by producing a counterexample to the putative luminosity of the condition of *feeling cold* – a condition that, to many at least,

---

condition on knowledge while objecting to the particular versions of it (putatively) needed for Williamson's argument.

<sup>4</sup> Williamson notes that some conditions are not susceptible to his anti-luminosity argument; such conditions might be *trivially* luminous. For example, conditions that never obtain are vacuously luminous, and conditions that always obtain might be luminous when presented under certain guises. Williamson's argument also does not work against eternal conditions, which always obtain if they ever obtain, though he gestures at an argument that shows that even such conditions are not plausibly luminous (2000, 108). Williamson's central point is not that there could be no luminous conditions, but rather that luminous conditions, if they exist, are "curiosities" (ibid., 109). Those conditions that we think of as paradigmatically luminous – e.g. being in pain, feeling cold, having a desire to  $\phi$ , its appearing that  $p$  – are, if Williamson is correct, non-luminous.

<sup>5</sup> That is, it is possible for her to introspect as assiduously as possible without thereby coming to know that she is in the condition.

seems paradigmatically luminous. Since analogous thought experiments can be produced for any other putatively luminous (non-trivial) condition, the anti-luminosity argument should generalise to all (non-trivial) conditions. Here is the thought experiment:<sup>6</sup>

**COLD MORNING:** S, a normal human, wakes up at dawn feeling freezing, very slowly warms up, and feels hot by noon. Throughout the morning S is concentrating sufficiently hard on the question of whether she feels cold, such that if she is in a position to know that she feels cold then she does indeed know. S's powers of discrimination are limited, and the change from S's feeling cold to hot is so gradual that S "is not aware of any change in them over one millisecond" (Williamson 2000, 7). S's confidence that she feels cold gradually diminishes, such that by noon she firmly believes that she no longer feels cold.

Let  $t_0, t_1, t_2 \dots t_n$  be a series of times at one-millisecond intervals from dawn to noon. Let  $\alpha_i$  be the case<sup>7</sup> at time  $t_i$ . Let C be the condition *that S feels cold*, and K(C) the condition *that S knows that C obtains*.

Now, let us assume that C is a luminous condition for S. That is, whenever S is in C, she is in a position to know she is in C. By the description of Cold Morning, whenever S is in a position to know that C obtains, she does in fact know that C obtains. Thus we have:

(LUM)        If C obtains in  $\alpha_i$  then K(C) obtains in  $\alpha_i$

Williamson then introduces the following margin-for-error principle, which he claims falls out of a simple safety condition on knowledge together with the description of Cold Morning:

(MAR)<sup>8</sup>        If K(C) obtains in  $\alpha_i$  then C obtains in  $\alpha_{i+r}$

---

<sup>6</sup> This is my description of the thought experiment, though it is similar to Williamson's original.

<sup>7</sup> A 'case' is a centred possible world – that is, a possible world with a designated subject and time.

By the description of Cold Morning, at dawn S feels cold, and at noon she no longer feels cold. So we have:

- (BEG)            C obtains in  $\alpha_0$   
(END)            C does not obtain in  $\alpha_n$

(LUM), (MAR), (BEG) and (END) are together incompatible. By (LUM), if C obtains in  $\alpha_0$ , then S knows that C obtains in  $\alpha_0$ . By (MAR), if S knows that C obtains in  $\alpha_0$ , then C obtains in  $\alpha_1$ . By (BEG), C does obtain in  $\alpha_0$ ; therefore, C obtains in  $\alpha_1$ . Similarly, we can establish that C also obtains in  $\alpha_2, \alpha_3, \alpha_4 \dots \alpha_n$ . So C obtains in  $\alpha_n$ . But according to (END) C doesn't obtain in  $\alpha_n$ . Thus we arrive at a contradiction.<sup>9</sup>

### 3. The problem with (MAR)

Since (BEG) and (END) simply follow from the description of Cold Morning, it seems we must either give up (MAR) or (LUM). Williamson counsels that we hold onto (MAR) and reject (LUM) – which is to say, abandon luminosity. Luminists, however, think that the lesson to be learned is that we should be suspicious of (MAR). But on what grounds?

Wong (2008) argues that (MAR) is derivable from the following two premises: (1) If in  $\alpha_i$  one knows that one feels cold, then in  $\alpha_i$  one feels cold, and (2) If in  $\alpha_i$  one feels cold, then in  $\alpha_{i+1}$  one feels cold. (1) follows uncontroversially from the

---

<sup>8</sup> There is no standard presentation of Williamson's argument in the literature, so (MAR), or its analogues in terms of possible worlds and times (rather than in the terminology of world-bound cases), is variously labelled (I) (Williamson 2000; Weatherson 2004; Blackson 2007; Ramachandran 2009); (R) (Vogel 2010); (C) (Wong 2008); (1) (Cohen 2010); and (KMAR) (Zardini forthcoming). I borrow (MAR) from Berker 2008.

<sup>9</sup> My presentation of Williamson's argument owes much to Berker 2008.

factivity of knowledge, but (2) is of course soritical. Thus Wong concludes that (MAR) is itself soritical and should be rejected in favour of (LUM). But Wong is wrong to think that this soritical argument is the only way to defend (MAR). In particular, Wong's defence of (MAR) does not appeal to a safety condition on knowledge, which Williamson clearly intends to be part of the justification for (MAR).

Weatherson (2004) and Vogel (2010) offer safety conditions on knowledge from which we can directly derive (MAR). According to both Weatherson's 'content safety' condition and Vogel's 'strong reliability' condition, one knows that a condition R obtains only if R obtains in all very similar cases. As Weatherson and Vogel argue, this version of the safety condition directly secures (MAR), but is itself implausible. Intuitively, for S's belief that R obtains to be sufficiently safe for knowledge, it must be that there are no sufficiently similar cases in which S has an untrue belief that R obtains.<sup>10</sup> But a more plausible understanding of the safety condition<sup>11</sup> – one that only mandates no nearby untrue belief – is insufficient to directly derive (MAR). For (MAR) states that it is a necessary condition on S's knowing that C obtains in  $\alpha_i$  that C also obtain in  $\alpha_{i+t}$ . But S's belief that she feels cold in  $\alpha_i$  could satisfy the 'no nearby untrue belief' condition so long as in all of the sufficiently similar not-C cases, S didn't *believe* that C obtained (cf. Berker 2008, 6). Luminists are right, then, to point out that a brute appeal to safety won't alone secure (MAR). And they are also right to point out that Williamson's anti-luminosity argument cannot run with safety alone; (MAR) is needed as a bridge principle from one moment to the next in order to deliver the *reductio* that S feels cold at noon. Thus the question is: how do we motivate (MAR)?

---

<sup>10</sup> I am assuming here and throughout my discussion of safety that for two cases of belief to be 'sufficiently similar' they must involve similar *methods* of belief-formation. Some specifications of safety factorise 'sufficiently similar' into a modal component and a methods or basing component.

<sup>11</sup> Like Weatherson's 'belief safety' condition and Vogel's 'moderate reliability' condition.

Williamson can be read as offering not one but two answers to that question, each relying on different specifications of the safety requirement on knowledge. The first defence employs a safety condition in terms of outright belief, which I call (BELIEF-SAFETY), and is intended to work on the assumption that there is no constitutive connection between one's feeling cold and believing oneself to feel cold. The second defence employs a safety condition in terms of degrees of confidence, which I call (CONFIDENCE-SAFETY), and is intended to allow for the possibility of such a connection. That is, the second defence is meant to secure (MAR), and thus the anti-luminosity argument, regardless of any possible constitutive connection that might obtain between the phenomenal and doxastic realms. My plan is to elaborate and bolster each of Williamson's defences of (MAR) in turn. In the first instance, I will argue that all we need to add to (BELIEF-SAFETY) to generate (MAR) is a plausible empirical supposition about the kind of creatures we are. In the second, I will counter accusations that Williamson's (CONFIDENCE-SAFETY) is an implausible condition on knowledge, and argue that, again, (MAR) can be derived from it together with a plausible assumption about our empirical character.

#### 4. Defending (MAR): non-constitutive accounts

Williamson's preliminary defence of (MAR) is intended to work on the assumption that there is no constitutive connection between the phenomenal and the doxastic – specifically, that one's feeling cold isn't constitutively tied to believing oneself to feel cold. This defence invokes a safety condition in terms of outright belief, which we can approximate as follows:

(BELIEF-SAFETY) In case  $\alpha$  S knows that a condition R obtains only if, in all sufficiently similar cases in which S believes that R obtains, it is true that R obtains

Roughly, (BELIEF-SAFETY) says that knowledge requires not just true belief, but the absence of nearby untrue belief.<sup>12</sup> (BELIEF-SAFETY) is intuitively plausible. Imagine I look through the window and form the true belief that it's raining outside. Unbeknownst to me, a prankster has placed a screen outside my window that projects an image of rain. Clearly I don't know it's raining outside, though it is. This is because in a nearby world – the world in which it has just stopped raining – I have the untrue belief that it's raining.

Despite the plausibility of (BELIEF-SAFETY), Vogel (2010) argues that we should reject it in favour of what he calls 'moderate reliability' (549), which amounts to this:

(VOGEL-SAFETY) In case  $\alpha$  S knows that a condition R obtains only if, in all sufficiently similar cases in which S believes that R obtains, it is not false that R obtains

The difference between (BELIEF-SAFETY) and (VOGEL-SAFETY) comes into play when (if ever) it is neither true nor false that R obtains.<sup>13</sup> Suppose that as S moves through Cold Morning, it is first true that S feels cold, then neither true nor false that S feels cold, and finally false that S feels cold. Now suppose that S is in

---

<sup>12</sup> What counts as a 'sufficiently similar case' (or a 'nearby world' or 'similar method') in definitions of safety is a vexed issue, analogous to what is known as the 'generality problem' for reliabilism (Conee and Feldman 1998). Williamson (2000) argues that the upshot of this problem is that we can offer no reductive analysis of reliability, and that our judgments about similarity of cases must be informed by our intuitive judgments about what constitutes an instance of knowledge or ignorance. This means that any claim to knowledge or ignorance is subject to dismissal via an alternative judgment about similarity of cases. This is equally true of the knowledge/ignorance claims involved in Williamson's anti-luminosity argument. That is, the luminist could simply argue that the possible not-cold case in which S believes she is cold is too dissimilar to the actual case  $\alpha$  to undermine the safety of S's belief in  $\alpha$ . In particular, the luminist could simply argue that S's possible false belief has a different basis from S's actual true belief (in the first case, S believes *on the basis of feeling cold* that she feels cold; in the latter case she believes on some other basis). These are both easy (if unconvincing) ways of defending luminosity against Williamson's putative counterexample. Cf. Weatherston 2004, 4.

<sup>13</sup> See Hawthorne 2005, Williamson 2005 and Zardini forthcoming for a discussion of the anti-luminosity argument and its relation to the phenomena of indeterminacy and vagueness.

the final instance of feeling cold, *t*. Can S know that she feels cold at *t*? (BELIEF-SAFETY) says no, while (VOGEL-SAFETY) says yes. Which safety condition gives us the correct treatment of cases such as these? Vogel offers the following thought experiment to motivate his version of safety:

UMPIRE: Imagine that there is an umpire who is invariably correct about every clear case of balls and strikes. That is, whenever television replay can discern one way or the other, the umpire is right, even on extremely close pitches. Every once in a while, the umpire calls as a ball a pitch that seems “too close to call” even on replay (549).

Vogel’s judgment is that the umpire is able to know about every clear case of balls and strikes, despite the fact in some of those cases he has nearby untrue beliefs. This result is compatible with (VOGEL-SAFETY) but not (BELIEF-SAFETY). My own feeling is that the umpire doesn’t know in all the clear cases, since whether or not he has any nearby false beliefs, he sometimes has nearby *mistaken* beliefs – that is, he sometimes believes pitches to be balls when they aren’t.<sup>14</sup> To my mind at least, the proximity of nearby beliefs that simply aren’t true seems sufficient to destroy knowledge. No doubt these issues are murky. In a footnote, Vogel offers another thought experiment:

COLOR CHIP: You see a number of color chips. Some are perfectly red, while the others are borderline red. The chips are placed in an urn, and one chip is chosen at random. Before you see the outcome, you believe that the selected chip will be red simpliciter, and it happens to be perfectly red (n. 15, 556).

Here, as Vogel himself agrees, it seems that you can’t know that the randomly chosen chip is red; this supports (BELIEF-SAFETY) over (VOGEL-SAFETY).

---

<sup>14</sup> The defender of (VOGEL-SAFETY) might protest, as an anonymous referee did, that it’s simply unfair to call such beliefs mistaken, since in these cases the proposition *that pitch is a ball* is neither true nor false. Might such cases not be a matter of ‘spoils to the victor’? Perhaps we bottom out at intuitions here. All I can say is that I have sympathy for the pitcher who walks a batsman on a pitch that is not in fact a ball, and laments that the umpire was mistaken. The pitcher would be wrong to think the umpire blameworthy for his call, but not wrong, I feel, to think the call mistaken.

It's also worth noting that (VOGEL-SAFETY) might have counterintuitive implications in cases of failed demonstratives.<sup>15</sup> Consider:

**PHANTOM TOLLBOOTH:** You see a tollbooth, point to it and say: "That's a tollbooth". However, you're on a drug that makes you hallucinate all sorts of things; in a very nearby world, the drug causes you to hallucinate a tollbooth.

Intuitively, you don't know that what you're pointing to is indeed a tollbooth. But in the nearby possible world in which you hallucinate a tollbooth, the demonstrative 'that' fails to refer to anything at all, since there is no tollbooth. In such a possible hallucinatory case it seems intuitive to say that your tollbooth-belief is neither true nor false.<sup>16</sup> This result is delivered by (BELIEF-SAFETY) but not (VOGEL-SAFETY). Finally, take a modalised version of Kripke's contingent liar cases:

**JACK AND JILL:** Jack believes that whatever Jill says next will be false. In a nearby world, the next thing Jill says is "whatever Jack believes is true".

Intuitively, Jack doesn't know that whatever Jill says next will be false, because in a nearby world that belief is not true – although that belief is not obviously false. Again, this intuition favours (BELIEF-SAFETY) over (VOGEL-SAFETY).

Let's assume that (BELIEF-SAFETY) is correct. Now recall:

(MAR) If  $K(C)$  obtains at  $\alpha_i$  then  $C$  obtains at  $\alpha_{i+1}$

To derive (MAR) from (BELIEF-SAFETY), what is needed is a principle that connects  $S$ 's belief about  $C$  in  $\alpha_i$  to  $S$ 's belief about  $C$  in  $\alpha_{i+1}$ . After all, asks

---

<sup>15</sup> For a discussion of how best to formulate safety in light of this sort of consideration, see Manley 2007.

<sup>16</sup> I'm assuming that the utterance "that's a tollbooth" expresses a belief in both the actual and nearby cases. This is not consistent with all accounts of singular thought.

Berker, might it not be possible for S to stop believing that she feels cold the precise moment she no longer feels cold? He writes:

[W]ho is to say that...as [S] gradually gets warmer and warmer during the course of the morning while carefully attending to how cold [she] feels, [S] stops feeling cold *before* [she] stops believing that [she] feels cold? (2008, 8).

If so, Berker suggests, then S's belief that she feels cold would satisfy (BELIEF-SAFETY) without (MAR)'s being true. To close the gap between (BELIEF-SAFETY) and (MAR), Berker proposes the following:

(BEL) If S believes C obtains in  $\alpha_i$ , then S believes C obtains in  $\alpha_{i+1}$

And indeed, (MAR) follows easily from (BELIEF-SAFETY) and (BEL). If S's believing she feels cold at one moment entails her believing she feels cold at the next, then by (BELIEF-SAFETY) if S *knows* she feels cold at one moment she must feel cold at the next. But (BEL) is untenable.<sup>17</sup> It is a soritical premise, since one can generate a paradox from it along with the assumptions that S believes that she feels cold at dawn and that S does not believe that she feels cold at noon (cf. Berker 2008, 7; Vogel 2010, 561). By the description of Cold Morning, S believes she feels cold at dawn and stops believing she feels cold sometime later; it cannot be true that belief in one case entails identical belief in the next.

No matter. Nothing as strong as (BEL) is needed to derive (MAR) from (BELIEF-SAFETY). According to (BELIEF-SAFETY), knowledge that one feels cold is incompatible with untrue belief that one feels cold in sufficiently similar cases. Now, suppose that in  $\alpha_i$  S feels cold, and in  $\alpha_{i+1}$  it is no longer the case that she feels cold. And suppose that in  $\alpha_i$  she believes truly that she feels cold and in  $\alpha_{i+1}$  she doesn't believe she feels cold. That is, her belief that she feels cold stops

---

<sup>17</sup> That is, understood as a universal generalisation. This applies to my discussion of (BEL\*) as well.

immediately with the cessation of her feeling cold, as per Berker's suggestion. Does S's belief that she feels cold in  $\alpha_i$  satisfy (BELIEF-SAFETY)? Not necessarily. For it could well be that in some sufficiently similar non-actual case  $\beta_{i+1}$ , S continues to believe she feels cold after she stops feeling cold.  $\alpha_i$  is a case in the actual world, as is  $\alpha_{i+1}$ . But there are also *non*-actual cases that are sufficiently similar to  $\alpha_i$  to destroy knowledge if S untruly believes in them that she is cold. To pass the safety test for knowledge, it is insufficient that one, as a matter of chance, lack untrue belief in all actual similar cases. One must also lack untrue belief in *possible* similar cases. This means that it is much easier for S to fail to know that she feels cold than some luminists seem to think. All that is required to derive (MAR) from (BELIEF-SAFETY) is something like the following:<sup>18</sup>

(BEL\*) If in case  $\alpha_i$  S believes she feels cold, then there exists a sufficiently similar possible case  $\beta_{i+1}$  in which S's cold-feelings are a phenomenal duplicate of her cold-feelings in  $\alpha_{i+1}$  and in which S believes she feels cold

Together, (BELIEF-SAFETY) and (BEL\*) yield (MAR). If in a given actual case, S knows that she feels cold, then by (BELIEF-SAFETY) there cannot be any very similar cases in which she believes she is cold but isn't. According to (BEL\*), if S believes she feels cold at one moment, there is some nearby world in which she believes she feels cold in the next moment (holding her feelings of cold at that moment fixed) (cf. Vogel 2010, 562). So, if in a given actual case S knows she feels cold, then she must feel cold in the next actual case – viz., (MAR).

---

<sup>18</sup> Berker suggests, but then rejects, a similar modalised version of (BEL), which he calls (BEL') (2008, 7, n. 11). I address Berker's objection to (BEL\*)/(BEL') shortly.

(BEL\*), unlike (BEL), is highly plausible.<sup>19</sup> First, (BEL\*) is just the kind of thing you would expect to be true of creatures like us. This is because we don't just believe at random. Our mental lives are structured by certain dispositions. When we believe something in one set of circumstances, in very similar circumstances we have a disposition to believe the same thing. (BEL\*) should be understood as encoding the empirical assumption that S, being a creature like us, shares these dispositions.<sup>20,21</sup> We might call this the *doxastic disposition premise*:

(DOXDIS)     If in condition R, S believes she is F, then for any condition R' very similar to R, S has some disposition in R' to believe she is F

(DOXDIS) seems to me fairly uncontroversial.<sup>22</sup> Imagine I am looking at a jar full of 1000 marbles; I don't know how many marbles there are in the jar, but I form the belief that there are a lot of marbles in the jar. If I'm then confronted with a

---

<sup>19</sup> Again, on the assumption that there is no constitutive connection between the phenomenal and the doxastic.

<sup>20</sup> Cohen (2010, 727-30) discusses deriving (MAR) from an empirical premise and a safety condition on knowledge (though Cohen's empirical premise repeats Williamson's talk of 'indiscriminability', which I find unhelpfully opaque). Cohen ultimately argues that this strategy does not work because it makes Williamson's argument circular (cf. Blackson 2010, 402):

[According to Williamson] our judgment that  $\alpha_{i+t}$  is similar to  $\alpha_i$  may require the judgment that if one could wrongly believe one feels cold in  $\alpha_{i+t}$ , then in  $\alpha_i$  one does not know one feels cold. Given [the empirical premise], this requires the judgment that in  $\alpha_i$  one knows one feels cold only if in  $\alpha_{i+t}$  one feels cold. And this is just the judgment that [(MAR)] is true (729).

Quite. Williamson's anti-luminosity argument requires that one make certain judgments about similarity, and on Williamson's view, these judgments are interdependent with our judgments about knowledge. Thus the anti-luminosity argument shows that if one wants to defend luminosity, one must deny that the cases in Cold Morning (and the possible phenomenologically identical cases) are similar, which seems strange. This might be circular, but it is only *viciously* so if the luminist were antecedently willing to deny what seems (to me at least) evidently true.

<sup>21</sup> There might be certain cases of possessing dispositions to believe  $p$  that do not entail having a nearby belief in  $p$ : for example, the disposition to believe of the Müller-Lyer illusion that the lines are of different length. Since I plausibly retain such a disposition even when it is well suppressed by my awareness of the illusion, this might be a case where a disposition to believe  $p$  doesn't entail a nearby belief in  $p$ . But such unusual cases are not my concern here.

<sup>22</sup> Though it might demand some refinement. Imagine that there is an extremely low objective chance that Hanna will believe  $p$  in condition R. Despite the terrible odds, she comes to believe  $p$  in R. Does she have a disposition to believe  $p$  in circumstances that are extremely similar to R? Perhaps not. But the issue needn't worry us here; I mean (DOXDIS) only to be able to handle the kind of central cases of belief-dispositions in play here.

very similar scenario – a jar with, say, 999 marbles, at a similar distance and in similar lighting conditions, etc. – I am disposed in that scenario to believe, again, that there are a lot of marbles in the jar. This is so even if I don't as it happens believe, in the second scenario, that there are a lot of marbles in the jar. Of course, what counts as a 'very similar condition' in (DOXDIS) will turn on, in part, what agents have the disposition to believe in the relevant situations. If I totally lack the disposition to believe the same in two situations,<sup>23</sup> this in part constitutes their not being very similar.

The luminist who wants to accept (BELIEF-SAFETY) while rejecting (MAR) is under pressure to deny a highly intuitive picture of how we work. In particular, he will have to maintain that, in the first case of not feeling cold, S has no disposition whatsoever to believe that she feels cold. It is insufficient, *pace* various luminists, that S simply happen not to believe that she feels cold when she stops feeling cold; she must lack even the disposition to so believe. But it is implausible that S – since she is, *ex hypothesi*, a creature like us – lacks such a disposition.<sup>24</sup> Thus it is implausible that any conditions of interest are, for us, luminous.

Second, (BEL\*), unlike (BEL), is not obviously a soritical premise. It does not trade on the vagueness of 'believes', but is instead a specific claim about what is true of S in Cold Morning; our assent to is secured by what we know about the doxastic dispositions of creatures like ourselves in situations like the one described in the thought experiment.

---

<sup>23</sup> Here and throughout I talk about having 'the disposition to believe the same in similar situations'. By this I mean not that for any two similar situations we will have the same disposition to believe in those situations, but rather that if in the first situation we believe something, then in the second we have a disposition to believe the same thing.

<sup>24</sup> Some luminists appear to think that what is at stake dialectically in the anti-luminosity debate is whether there are any possible creatures who enjoy luminosity, not whether we are such creatures ourselves. See section 5 for a discussion of why this is misguided.

While Berker acknowledges that (BEL\*) is not straightforwardly soritical in the way that (BEL) is, he argues that (BEL\*) nonetheless has a soritical consequence. This is because repeated applications of (BEL\*) yield the conclusion that it is possible for S to feel extremely hot while believing she feels cold.<sup>25</sup> Berker claims that this is unacceptably absurd.<sup>26</sup> He writes:

I think we should have serious doubts that such a case is even possible – serious doubts that there could exist a being who counts as having beliefs and experiences, and yet whose beliefs and experiences are as wildly at odds with one another as they would be in [the case in which one feels extremely hot but believes oneself to feel cold] (2008, n. 11, 7-8,).

Is it really so hard to countenance such a possibility? The similarity relation is intransitive, so a case in which one felt extremely hot but believed oneself to feel cold would be a case very dissimilar to the one imagined in *Cold Morning*. In particular, the intransitivity of ‘very similar method’ means that, in such a case, one might very well be using a method very different from the one a normally functioning person uses to form beliefs about her feelings of cold. One could, for example, be the victim of prolonged psychological priming, or in the grip of a certain philosophical picture of the mind that makes one systematically distrust one’s inclinations to judge one’s own phenomenal state. Is it really so hard to imagine someone in these conditions coming to believe she feels cold when she actually feels extremely hot? These possible cases might be remote, no doubt. But their existence – like the existence of bad sceptical worlds – does nothing to undermine S’s ability to know in normal situations. That (BEL\*) implies that they are possible is thus no knock against it. In any case, as Berker himself notes, this objection to (BEL\*) seems motivated by a view on which the phenomenal and the

---

<sup>25</sup> Berker in fact claims that repeated applications of (BEL\*) yield the conclusion that one could believe oneself to feel cold while feeling as hot as “if one were in the center of the sun” (n. 11, 7). Since it’s doubtful that one would feel much of anything if one were in the centre of the sun, I take it that Berker just means ‘extremely hot’. It’s worth noting that one can’t generate this consequence from repeated applications of (BEL\*) alone – one would need analogous principles that apply not just to  $\alpha$  cases, but also  $\beta$  cases and so forth. It’s also worth noting that if there is an upper bound to how cold or hot one can feel, then not all of these analogous principles will be true.

<sup>26</sup> Thanks to an anonymous referee for pointing this out to me.

doxastic enjoy a constitutive connection. Such a view is not my target in this section, and (BEL\*) will not feature in my eventual argument against it.

Before moving to my second defence of (MAR), one that dispenses with the assumption that there is no constitutive connection between the phenomenal and the doxastic, let me take quick stock. On my favoured reconstruction of the anti-luminosity argument (for non-constitutive accounts), its essence is this. Suppose that S, like us, is the sort of creature for whom believing something in one situation means having, in extremely similar situations, the disposition to believe the same thing. And suppose that S is in a condition C, but in what we might call a *liminal* case of it.<sup>27</sup> That is, there is an extremely similar case to the one she is actually in which is not a case of C. Imagine that S believes she is in C; is this belief knowledge? It seems not. For in the very similar non-C case, S has the disposition to believe she is in C. Thus, her true belief that she is in C is rendered unreliable by a nearby untrue belief that she is in C. So in such a liminal case of being in C, S cannot know that she is in C. This argument generalises to all non-trivial conditions and for all subjects whose beliefs are structured by these sorts of dispositions. That is, it generalises to all interesting mental state conditions in which creatures like us plausibly find ourselves.

My reconstruction of the anti-luminosity argument differs from the original in avoiding Williamson's favoured talk of our 'limited powers of discrimination'. According to Williamson, it is this cognitive feature that drives the anti-luminosity argument (2000, 12, 97, 103-4). While Williamson unpacks the idea in various ways,<sup>28</sup> I take its essence to be this.<sup>29</sup> When we are thinking about whether or not

---

<sup>27</sup> The appeal to liminal cases here is not an appeal to the vagueness of the concept of C. One can imagine the borders firmly fixed. So, for example, imagine that S is in condition C just in case she is experiencing at least 100 units of x (where x is some phenomenal experience, e.g. feelings of cold), and imagine that she is experiencing exactly 100 units of x. This would be a liminal case of S's being in C. I avoid the more obvious terminology of 'borderline' cases because of its associations with semantic/conceptual vagueness.

<sup>28</sup> Some of which can sound somewhat question-begging. Consider the following:

we are in some sort of state, we turn our attention to the relevant underlying phenomenon that constitutes that state. When it comes to figuring out whether there is any milk in the fridge, we train our attention on the contents of the fridge. Similarly, when it comes to figuring out whether we feel cold, we turn our attention to our sensations of cold.<sup>30</sup> Now, it is a disappointing truth of our perceptual capacities that they are not infinitely discriminating: we cannot always tell of two distinct things whether they are indeed distinct. If, by chance, we come to believe that two indiscriminable situations are in fact different, this lucky belief does not constitute knowledge. Thus, our limited perceptual capacities limit our ability to know. Williamson's anti-luminosity argument can be understood as an application of this observation to the phenomenal sphere. In training our attention on our underlying sensations of cold and hot, we cannot reliably distinguish between two extremely similar sensations, one of which is cold and the other of which is not. Since reliability is required for knowledge, we cannot know that we are cold in such a case.<sup>31</sup> Of course, the luminist can simply reply that this begs the question against him. For any defender of the luminous could simply reject the idea that our powers of discrimination *are* limited. But this is to maintain not only that we have privileged access to our mental states, but also that our perceptual

---

The main idea behind the argument against luminosity is that our powers of discrimination are limited. If we are in a case  $\alpha$ , and a case  $\alpha'$  is close enough to  $\alpha$ , then for all we know we are in  $\alpha'$ . Thus what we are in a position to know in  $\alpha$  is still true in  $\alpha'$ . Consequently, a luminous condition obtains in  $\alpha$  only if it also obtains in  $\alpha'$ , for it obtains in  $\alpha$  only if we are in a position to know that it obtains in  $\alpha$  (2000, 12).

One might have a similar worry about Williamson's stipulation that "[S] is not aware of any change in [her feelings of cold or hot] over one millisecond" (ibid., 97) and that she is "almost equally confident that [she] feels cold, by the description of the case" (ibid., 97). These glosses on the argument can sound more like re-descriptions of its conclusion than reason to accept it.

<sup>29</sup> Cf. Vogel (2010, part II) for a discussion of what Williamson might mean by 'limited powers of discrimination', and how this generally relates to the luminosity doctrine.

<sup>30</sup> This presupposes a broadly perceptual model of self-knowledge. For a discussion of alternative models and how they interact with the anti-luminosity argument, see section 5.

<sup>31</sup> It's worth noting that the mere ability to distinguish dissimilar things is not the same thing as the ability to correctly categorise them under concepts. Thus I might be able to distinguish two distinct phenomenal sensations without being in a position to know that one is a feeling of cold and another isn't. (Compare: I might be able to distinguish between the sound of French and the sound of German, but this doesn't mean that I'm always in a position to know that I'm hearing French rather than German). If so, then our limited powers of discrimination are not, *pace* Williamson's suggestion, the only obstacle to the luminosity of the mental.

capacities when it comes to attending to those mental states, unlike attending to external world states, are infinitely discriminating. And this might seem like a large bullet to bite.

How does Williamson's original argument (on my interpretation) square with my own preferred reconstruction? My empirical premise – (DOXDIS) – is certainly compatible with Williamson's (also empirical) claim that our powers of discrimination are limited. But it is also compatible with a variety of other stories we might want to tell about how our phenomenal beliefs arise.<sup>32</sup> Whatever version of that story we embrace, it seems undeniable that creatures like us are disposed to believe the same things in extremely similar situations. Of course, what we mean by 'similar' here matters crucially. The luminist can always resist the claim, necessary for (BEL\*) to follow from (DOXDIS), that the situations from one moment to the next in *Cold Morning* are, indeed, 'very similar'. And, as Williamson himself admits, our judgments about similarity are inevitably bound up with our epistemic judgments (2000, 100-1). So the luminist can always dig in his heels somewhere; my formulation of the argument in terms of doxastic dispositions will not compel the dogmatic luminist to abandon his views. But, if my reconstruction of Williamson's argument works, then the price the luminist has to pay for resisting it is very high indeed. He must either deny that we are creatures who are disposed to believe the same in similar situations, or insist that the cases from one moment to the next in Williamson's thought experiment should not count as similar. Both options seem to me perverse. As such, defending (MAR) by invoking a safety condition on knowledge together with the minimal (DOXDIS) seems to me dialectically preferable to Williamson's own defence in terms of finite powers of discrimination.

---

<sup>32</sup> These alternative stories become particularly important when we consider the possibility, discussed in the next section, that the phenomenal is constitutively connected to the doxastic.

## 5. Defending (MAR): constitutive accounts

In defending (MAR) against the luminists, I argued that it derives easily from a simple safety condition on knowledge, (BELIEF-SAFETY), coupled with the premise (BEL\*), which in turn is justified by the empirical premise (DOXDIS). However, (BEL\*) is subject to complaint from those luminists who maintain that phenomenal conditions such as feeling cold enjoy a constitutive connection to one's beliefs about whether one feels cold. As such, the anti-luminosity argument as elaborated so far is impotent against a popular view of the phenomenal (Weatherson 2004; Berker 2008; Ramachandran 2009).<sup>33</sup> Suppose that the following were true of the relationship between S's beliefs about feeling cold and the facts about her feeling cold:

(CON) If S has done everything she can to decide whether she feels cold, then she believes that she feels cold if and only if she feels cold<sup>34</sup>

Since by the description of Cold Morning, S is doing everything she can to decide whether she feels cold, her coming to believe that she feels cold is both necessary and sufficient for her indeed feeling cold. Now recall (BEL\*):

(BEL\*) If in  $\alpha_i$  S believes she feels cold, then there exists a sufficiently similar possible case  $\beta_{i+1}$  in which S's cold-feelings are a phenomenal duplicate of her cold-feelings in  $\alpha_{i+1}$  and in which S believes she feels cold

---

<sup>33</sup> Ramachandran proposes that feeling cold might be a response-dependent condition, while Weatherson argues along physicalist lines that feeling cold and believing oneself to feel cold could in fact consist in the same brain state. There are various possible ways of unpacking what the constitutive connection between feeling cold and believing oneself to feel cold might amount to.

<sup>34</sup> Adapted from Berker 2008, 9.

If (CON) is true, then (BEL\*) is false. Why? By the description of Cold Morning, there is some value  $i$  such that S believes she feels cold in  $\alpha_i$  and no longer believes she feels cold in  $\alpha_{i+1}$ . (CON) entails that for that value of  $i$ , S does indeed feel cold at  $\alpha_i$  and no longer feels cold  $\alpha_{i+1}$ . And according to (CON), any case that is a phenomenal duplicate of  $\alpha_{i+1}$  (with regard to S's feelings of cold) will also be a *doxastic* duplicate of  $\alpha_{i+1}$ . So (BEL\*) goes false for the value of  $i$  such that S believes she feels cold in  $\alpha_i$  and no longer believes she feels cold in  $\alpha_{i+1}$ . For that value of  $i$ , it is not true that there exists a possible case  $\beta_{i+1}$  in which S's cold-feelings are a phenomenal duplicate of her cold feelings in  $\alpha_{i+1}$  and in which S believes C obtains. Thus, if (CON) is true, (BEL\*) is false.

To defend (MAR) without assuming, as we have been doing, that (CON) is false, we need to appeal to Williamson's refined safety requirement in terms of degrees of confidence.<sup>35,36</sup> We might specify this safety condition as follows:

(CONFIDENCE-SAFETY) If in case  $\alpha$  S knows with degree of confidence  $c$  that she is in a condition R, then in any sufficiently

---

<sup>35</sup> Williamson distinguishes these from degrees of subjective probability measured by one's betting behaviour. A degree of confidence is a degree of *outright belief*.

Intuitively, one believes  $p$  outright when one is willing to use  $p$  as a premise in practical reasoning. Thus one may assign  $p$  a high subjective probability without believing  $p$  outright, if the corresponding premise in one's practical reasoning is just that  $p$  is highly probable on one's evidence, not  $p$  itself...we can think of one's degree of outright belief in  $p$  as the degree to which one relies on  $p$ . Outright belief in a false proposition makes for unreliability because it is reliance on a falsehood (2000, 99).

On Williamson's view, one can have a certain degree of confidence (outright belief) without having *an* outright belief. This non-standard distinction has created a lot of confusion about Williamson's argument (e.g. Blackson 2007). Cf. Ramachandran's discussion (2009, 663, especially n. 3). Ramachandran's first interpretation (he offers four in total) of Williamson's argument relies on a (deliberate) misreading of Williamson's notion of confidence.

<sup>36</sup> Cohen (2010) argues, quite correctly, that (CONFIDENCE-SAFETY) cannot be derived from (BELIEF-SAFETY) without a soritical premise, namely that a slight change in degree of confidence does not affect whether one believes outright. He thus concludes that (CONFIDENCE-SAFETY) cannot be generally motivated. However, Williamson intends (CONFIDENCE-SAFETY) to be his fully elaborated safety condition, itself an intuitively plausible gloss on the reliability requirement for knowledge. Here I defend it as such.

similar case  $\alpha^*$  in which S has an at-most-slightly-lower degree of confidence  $c^*$  that she is in condition R, it is true that she is in R

The idea behind (CONFIDENCE-SAFETY) is that for one to know that one is in a given condition, it cannot be the case that one is almost as confident that one is in that condition – *even if* that confidence is short of full-fledged belief – in a sufficiently similar situation. That is, nearby misplaced confidence – high confidence in an untruth – is sufficient to preclude knowledge. Let us grant for the sake of argument that S’s feeling cold is constitutively connected to her belief that she feels cold in the way (CON) specifies. Suppose that in  $\alpha_i$  S truly believes that she feels cold, and that in  $\alpha_{i+1}$  she is still quite confident that she feels cold, but insufficiently confident for outright belief. By (CON), S feels cold in  $\alpha_i$  but does not feel cold in  $\alpha_{i+1}$ . But by (CONFIDENCE-SAFETY), S does not know that she feels cold in  $\alpha_i$ . So even a constitutive connection between one’s feeling cold and believing oneself to feel cold is insufficient to vindicate luminosity.

Again, to derive (MAR) from (CONFIDENCE-SAFETY) one needs an additional assumption that links S’s confidence that she feels cold to her confidence that she feels cold in nearby cases. Berker proposes something like the following:

(CONF)        If in  $\alpha_i$  S has degree of confidence  $c$  that she feels cold, then in  $\alpha_{i+1}$  S has an at-most-slightly-lower degree of confidence  $c^*$  that she feels cold

While Berker does not wish to dispute (CONF) – saying that it “seems indisputable, given the description of the situation at hand” (2008, 12) – it is again worth noting that, as with (BEL), (CONF) is unnecessarily strong. To generate (MAR) from (CONFIDENCE-SAFETY), the following weaker premise will do:

(CONF\*) If in  $\alpha_i$  S has degree of confidence  $c$  that she feels cold, there exists a sufficiently similar possible case  $\beta_{i+1}$  in which S's cold-feelings are a phenomenal duplicate of her cold-feelings in  $\alpha_{i+1}$  and in which S has an at-most-slightly-lower degree of confidence  $c^*$  that she feels cold

It should be clear that (CONF\*) is even more plausible than (CONF), again by appeal to an empirical premise:

(DOXDIS\*) If in condition R, S believes with confidence level  $x$  that she is F, then for any condition R' very similar to R, S has some disposition in R' to believe with confidence levels similar to  $x$  that she is F

Roughly, (DOXDIS\*) is the claim that if in a certain situation S has a certain confidence level, then in a very similar situation she is disposed to have a very similar confidence level.<sup>37</sup> As such, if in  $\alpha_i$  S has degree of confidence  $c$  that she feels cold, there exists a sufficiently similar possible case in which S feels just as cold as she does in  $\alpha_{i+1}$  and has a degree of confidence  $c^*$  that is at most slightly lower than  $c$  that she feels cold. Like (BEL\*), (CONF\*) is not a soritical premise, since it cannot be used to generate the (obviously false) conclusion that S has the same or similar degree of confidence at dawn and noon that she is in C. Rather than trading on the vagueness of any of its constituent terms, (CONF\*) encodes a plausible empirical premise about our dispositions to believe similarly in similar situations. And finally, unlike (BEL\*), (CONF\*) is compatible with (CON). While (BEL\*) entailed that S's belief about whether she feels cold could possibly come

---

<sup>37</sup> (DOXDIS) and (DOXDIS\*) can be subsumed under a more general empirical principle: If in condition R, S has attitude A towards  $p$ , then for any similar attitude A' and similar condition R', S in R' has some disposition to have A' toward  $p$ .

apart from the fact of whether she feels cold – in contradiction with (CON) – (CONF\*) doesn't entail anything of the sort. Instead, (CONF\*) merely entails that S's confidence levels about her feelings of cold are similar in similar cases. This is perfectly compatible with the view that S's believing she feels cold is both necessary and sufficient for her feeling cold.

In any case, it is at (CONFIDENCE-SAFETY) that many luminists direct their ire (Leitgeb 2002; Berker 2008; Ramachandran 2009; Cohen 2010). They hope to show that (CONFIDENCE-SAFETY) is a non-genuine condition on knowledge, leaving us only with the original defence of (MAR) in terms of (BELIEF-SAFETY), and thus the constitutive connection view of the phenomenal unscathed by the anti-luminosity argument. For example, in his attack on (CONFIDENCE-SAFETY), Cohen argues:

[I]t is not obvious why one's confidence at  $t_i$  is misplaced. We are supposing that at  $t_i$  one knows one feels cold. Thus at  $t_i$  one feels cold and one believes one feels cold. It follows that if at  $t_{i+1}$  one no longer believes one feels cold, then at  $t_i$  one just barely believes one feels cold. Now suppose one no longer feels cold at  $t_{i+1}$ . Then one just barely feels cold at  $t_i$ . So under these suppositions, at  $t_i$  one just barely feels cold and one just barely believes one feels cold. So how is one's confidence at  $t_i$  misplaced? (2010, 726).

In a similar spirit, Berker writes:

[W]hy should we withhold the honorific 'reliable' in the kinds of cases Williamson describes? ... [W]hat if one's belief that  $p$  tapers off (as it were) just as its being the case that  $p$  tapers off, and in precisely the same way?... [(CONFIDENCE-SAFETY)] deems as unreliable belief-forming mechanisms that appear to be as reliable as they could possibly be (2008, 12).

Spelling out this line of objection, Berker proposes that we think of S's feelings of cold in terms of numbers of "freezons". He imagines that at dawn S experiences 50 freezons worth of cold, that at noon she experiences -50 freezons worth of cold, and at any time during the day her degree of confidence that she feels cold is

directly correlated with her subjective feeling of cold as measured in freezons. Finally, Berker supposes that S believes she feels cold if and only if she indeed feels cold. That is, S's confidence that she feels cold drops below the belief-threshold at precisely the same moment she ceases feeling cold. If (CONFIDENCE-SAFETY) is true, then S, at some point during the day, fails to know that she feels cold. But Berker objects that this is absurd, since "[S's] beliefs about whether [S] feels cold appear to be as reliable as they possibly could be" (2008, 13). In the same vein, Ramachandran writes:

[(CONFIDENCE-SAFETY)] is too strong a requirement because it would rule out luminosity in the hypothesized 'perfect-calibration' situation, which is daft, because one *couldn't* be any more reliable (2009, 668).<sup>38,39</sup>

While there is something no doubt attractive in the Leitgeb-Berker-Ramachandran-Cohen thought that this 'perfect-calibration situation' represents some sort of maximally reliable possibility – and thus that (CONFIDENCE-SAFETY) can't possibly be a genuine condition on knowledge – that attraction isn't too difficult to dispel. Consider the following case, analogous to Berker's freezons case:

**GLASS HALF FULL:** Henry likes watching empty glasses slowly fill with water until they are full. In normal conditions and when he is paying close attention, Henry's confidence that a given glass is at least half full is directly correlated with how full the glass is, rising slowly from 0% to 100% confidence as the initially empty glass is slowly filled to the brim. Moreover, Henry believes that glasses are at least half full if and only if they are indeed at least half full. It thus follows that the confidence threshold for outright belief is 50%. The only proposition Henry entertains about a glass as it fills is *that the glass is at least half full*.

---

<sup>38</sup> See also Leitgeb 2002, 216.

<sup>39</sup> Ramachandran (668) ultimately does not endorse this line of argument on the grounds that the individual it imagines is too different from the subject in Cold Morning. I take this to be a bad objection because the issue is whether (CONFIDENCE-SAFETY) is a genuine condition on knowledge in general.

Does Glass Half Full, as Berker et al. suggest, represent a ‘perfect-calibration situation’? Surely not. When the glass is only a fifth full – that is, very obviously less than half full, even to an average estimator – Henry still has a 20% confidence that the glass is at least half full. And when the glass is four-fifths full – that is, very obviously more than half full, even to an average estimator – Henry has only an 80% confidence that the glass is at least half full. Henry’s confidence profile is hardly a maximally reliable possibility. The suggestion that Henry, or Berker’s freezons subject, represents a ‘perfect-calibration situation’ is thus misguided. So it cannot be reason to think that either case constitutes a counterexample to (CONFIDENCE-SAFETY).

For a more plausible counterexample to (CONFIDENCE-SAFETY), take the following case:

**GLASS HALF EMPTY:** Henrietta likes watching full glasses slowly drain until they are empty. The confidence threshold for outright belief is 80%, and, in normal conditions and when she is paying close attention, Henrietta believes of a given glass that it is at least half empty if and only if it is indeed at least half empty. Henrietta’s confidence that the glass is at least half empty remains near 0% until the glass is almost half empty, then steeply increases to 80% just as the glass is exactly half empty, increasing to and remaining at 100% shortly thereafter. The only proposition that Henrietta entertains as the glass empties is *q*, *that the glass is at least half empty*.

Henrietta, unlike Henry, is a very good judge of the fullness or emptiness of glasses. For Henrietta doesn’t have relatively high confidences in *q* when it is obviously false (e.g. when the glass is only a fifth empty), or relatively low confidences in *q* when it is obviously true (e.g. when the glass is four-fifths empty). Since all of Henrietta’s *q*-beliefs are reliably true, they satisfy (BELIEF-SAFETY).<sup>40</sup> Nonetheless, not all of Henrietta’s *q*-beliefs satisfy (CONFIDENCE-

---

<sup>40</sup> So do Henry’s glass-beliefs, of course, since he believes a glass to be at least half full if and only if it is indeed so. So what I say about Henrietta goes for Henry as well; I’ve only introduced the Glass Half Empty case to clear away the distractions that arise from Henry’s putatively ‘perfect’ calibration.

SAFETY). Specifically, Henrietta's belief in  $q$  when it is exactly half empty does not satisfy it. For just a moment earlier – when the glass was *almost* half empty – Henrietta had a high confidence just short of belief that the glass was at least half empty. If it seems plausible that Henrietta knows that the glass is at least half empty when it is exactly half empty, this tells against (CONFIDENCE-SAFETY) in favour of the less demanding (BELIEF-SAFETY).

The crucial question then is: does Henrietta know that the glass is at least half empty when the glass is exactly half empty? There is reason to think not. After all, even if Henrietta reliably tips over the threshold to outright belief just when it becomes true that the glass is at least half empty, it's also the case that Henrietta reliably *almost* believes that the glass is at least half empty when it's not. That is, in nearby cases in which it is false that the glass is at least half empty, Henrietta has a high confidence that it is true. Specifically, at a point at which the glass is just less than half empty, Henrietta is 79% confident that the glass is at least half empty: very confident indeed. Does this not undermine Henrietta's claim to know  $q$  just a moment later, when the glass is exactly half empty?

A common picture of the relationship between confidence and practical reasoning would suggest it does. According to that picture, one's confidence in a proposition  $p$  is a measure of one's willingness to rely on  $p$  as a premise in practical reasoning; one believes outright when one's willingness to rely on  $p$  crosses a certain threshold.<sup>41</sup> If so, then a confidence just short of outright belief in  $p$  will yield some cases in which one uses  $p$  in one's practical reasoning despite not believing  $p$  outright. When the glass is only, say, 49% empty, Henrietta's 79% confidence that the glass is at least half empty means that she has some tendency to use that false proposition as a premise in her practical decision-making. If she were making several  $q$ -relevant but independent decisions at the same time, we could expect to

---

<sup>41</sup> That is to say, one's believing  $p$  isn't anything over and above meeting a certain threshold for confidence – e.g. putting  $p$  in one's belief-box or mentally asserting  $p$ .

see Henrietta acting on  $q$  when it is false. This nearby willingness to act on  $q$  when it is untrue plausibly undermines Henrietta's claim to know  $q$  in the case where  $q$  has just become true.

Perhaps though the most compelling case in favour of (CONFIDENCE-SAFETY) can be made by tweaking the thought experiments that motivate safety in the first place. Take for example:

RECEDING FAKE BARNES: Mirra is looking at two rows of what look like barns in the distance. The first row is made up of real barns; the second row is fake. In situations like this, Mirra only forms beliefs about the proposition *that is a row of barns*, and she reliably forms only true beliefs about that proposition. The threshold for outright belief is 70% confidence. Of the first row, Mirra believes with 70% confidence that it is a row of barns. Of the second row, Mirra believes with 69% confidence that it is a row of barns.

Mirra's belief that the first row is a row of barns is reliably true; she could not have easily had the untrue belief that it is a row of barns. But it nonetheless seems somewhat odd to say that Mirra knows that the first row is a row of barns. After all, Mirra has a 69% confidence that the fake barns right behind the real barns constitute a row of barns. This suggests that safety requires more than the absence of nearby untrue belief; it requires the absence of nearby untrue *almost*-belief. If so, then luminosity requires not only that our phenomenal beliefs satisfy (BELIEF-SAFETY); it requires that our phenomenal beliefs further satisfy (CONFIDENCE-SAFETY). Since the former but not the latter can be satisfied by beliefs that enjoy a constitutive connection to their phenomenal contents, it seems that no non-trivial mental conditions are luminous.

This is not to say that *no* possible creature could have a degree-of-confidence profile that satisfied both (CONFIDENCE-SAFETY) and luminosity. Take again Henrietta, who believes  $q$ , that the glass is at least half empty, if and only if  $q$  is true. Imagine that at the point at which  $q$  becomes true, Henrietta's confidence

level discontinuously jumps from 0% to 80%. In so doing, Henrietta believes not just truly but moreover *safely* that the glass is at least half empty whenever it is indeed at least half empty. Thus for her the condition of the glass being at least half empty is luminous. Analogously, if the subject of Berker's freezon case were to exhibit a jump down from a high to low confidence at the point at which she no longer feels cold,<sup>42</sup> then for her, feeling cold would be luminous. As Berker himself notes, physical systems are rarely characterised by such discontinuity (2008, 15). As such, it's very implausible that for creatures like us anything very much is luminous.

Berker however draws a different lesson. He claims that to maintain that (CONFIDENCE-SAFETY) is a genuine condition on knowledge is to "insist not just that *reliability* is required for knowledge, but moreover that *perfect reliability* is required, and that way scepticism lies" (ibid., 15). But this is misleading. It is the objector to (CONFIDENCE-SAFETY), not its proponent, who demands perfection. Even Henry, whom all will admit is a pretty dismal estimator, gets to know some of the time, according to (CONFIDENCE-SAFETY), that the glass is at least half full. And similarly for Berker's freezon subject. (CONFIDENCE-SAFETY) will deliver even more generous verdicts when it comes to imperfect epistemic agents like us; while diverging quite a bit from the maximally reliable confidence profile, we still get to know much of the time that we are in various conditions. (CONFIDENCE-SAFETY) doesn't generally demand perfection: far from it. Rather, it implies that we are not perfect, which is a different thing altogether. Our implied imperfection is no reason to reject (CONFIDENCE-SAFETY). Or, rather, it is no reason beyond an antecedent affection for luminosity.

---

<sup>42</sup> Berker describes such a confidence profile (see his Figure 2) as one in which one is absolutely certain (i.e. has confidence 1) that one feels cold when one indeed feels cold, and drops off to confidence 0 when one doesn't feel cold. But nothing this extreme is required. (CONFIDENCE-SAFETY) and the luminosity of feeling cold are together consistent with a more leisurely descent in confidence both above and below the belief-threshold; all that is required is a discontinuous drop-off at the point where one ceases to feel cold (in the actual and very similar cases).

It's worth noting that the luminist could simply deny that (CONFIDENCE-SAFETY) is a genuine condition on knowledge of the phenomenal, even if he accepts it as a genuine condition on knowledge of external states like the fullness of a glass or the presence of barns. His willingness to do so will turn on the particular picture of the phenomenal-doxastic connection to which he is committed, or willing to commit himself in order to vindicate luminosity. If, for example, one maintained that feeling cold was simply a matter of believing oneself to feel cold after deliberation – that there was no underlying phenomenon of 'feeling cold' that our beliefs track or fail to track – then one could plausibly insist that models of reliability that are based on perceptual capacities are inapplicable. Thought experiments like Glass Half Empty and Receding Fake Barns would then be orthogonal to the issue. In this way, a commitment to (CON) might still offer a weapon of resistance against (MAR), and thus the anti-luminosity argument. But for reasons I discuss in the next section, the weapon might be double-edged.

## 6. Conclusion: The Dialectical Situation

Let me take stock. Luminists typically argue that (MAR) cannot be derived from plausible premises: either it requires an overly strong safety condition, or it requires additional soritical premises. I've offered a reconstruction of Williamson's two-part defence of (MAR), the first part intended to work on the assumption that there is no constitutive connection between the doxastic and the phenomenal, and the second intended to work without that assumption. In each case I've argued that an independently plausible safety condition – (BELIEF-SAFETY) and (CONFIDENCE-SAFETY), respectively – combined with an independently plausible view of our empirical characters, yields (MAR). In brief, our coarse-grained dispositions to believe renders us incapable of knowing that we are in liminal cases of (otherwise luminous) conditions. This is plausibly true whether or

not there exists a constitutive connection between the phenomenal and the doxastic, at least assuming that such a connection preserves the status of self-knowledge as an instance of knowledge more generally. Thus the luminist who wants to dig in his heels either must reject a plausible picture of our empirical character, or isolate self-knowledge as a *sui generis* epistemic category. I want to end by saying something about these two possible retreats for the luminist, and how they figure in the overall dialectical situation at hand.

The first possible retreat for the luminist is to reject what I claim is a plausible empirical characterisation of the kinds of creatures we are, namely creatures who have certain coarse-grained dispositions to believe. But some luminists, in arguing against Williamson, seem to think that it is sufficient to show merely that *some* metaphysically possible creature could have a different empirical character, and so enjoy luminosity (Berker 2008; cf. Weatherson 2004, 9). Williamson himself is not maximally explicit about the conditions to which his anti-luminosity argument applies: conditions that human and human-like creatures actually find themselves in, conditions that are metaphysically possible, conditions that are logically possible?<sup>43</sup> But to think the anti-luminosity debate is about possible creatures, not creatures like us, is to mistake the dialectical situation. If the anti-luminosity argument were meant to apply to all possible creatures, we need nothing very much to show it to be impotent. For the anti-luminosity argument is trivially compatible with various possible creatures for whom feeling cold and a whole range of other conditions are luminous.<sup>44</sup> More significantly, the question of whether we are luminous is the question we *should* care about. The possibility of creatures, perhaps radically different from ourselves, for whom interesting conditions are luminous does little to assuage the live possibility that our

---

<sup>43</sup> Somewhat gnomically, Williamson writes, prefacing the anti-luminosity argument: “The domain of cases will be taken to include counterfactual as well as actual possibilities. Since the cases on which the arguments below rely are physically and psychologically feasible, issues about the bounds of possibility are not pressing” (2000, 94).

<sup>44</sup> See n. 12.

philosophy of mind, epistemology, and ethics are all built on disreputable Cartesian foundations.

The second possible retreat for the luminist is to endorse a particular kind of constitutive connection view. While a variety of non-perceptual accounts of self-knowledge are compatible with there being a constitutive connection between the phenomenal and the doxastic,<sup>45</sup> according to one such popular account, self-knowledge is deeply disanalogous with knowledge of the external world. That we are always in a position to know our own phenomenal states is a conceptual truth, or a feature of our ‘grammar’ in the Wittgensteinian sense of that term (Shoemaker 1986; Wright 1989, 1998; Heal 2001; Bilgrami 2006; Coliva 2009). Self-knowledge is sometimes thus said to be ‘no cognitive achievement’. Such a view isolates self-knowledge as a *sui generis* epistemic state, thereby freeing it from the normal requirements for knowledge like reliability or truth-tracking. Obviously, whether this offers a legitimate retreat for the luminist depends on whether the phenomenal and the doxastic are really connected in just this way – whether self-knowledge is really just a conceptual or grammatical upshot. Again, this is in part an empirical question about the kinds of creatures we are. I don’t mean to enter this debate. But I will close with two observations about it.

The first is that those philosophers of mind who endorse constitutive accounts of self-knowledge are often motivated to do so because they wish to vindicate the luminosity of the mental. Thus Berker misrepresents the dialectical situation somewhat when he claims that

typically it is *precisely because* they think that there is a tight connection between certain mental states and beliefs about those states that some philosophers claim the mental states in question to be luminous (2008, 9).<sup>46</sup>

---

<sup>45</sup> See e.g. Chalmers 2003 and various essays in Coliva 2012.

<sup>46</sup> See also Ramachandran 2009, 666.

That some philosophers favour constitutive accounts does not make luminosity more plausible, since it is for the vindication of luminosity that these accounts were generally designed.<sup>47</sup> My second point is this. The tension between Williamson's anti-luminosity argument and the 'no cognitive achievement' view of self-knowledge is deep and fundamental. On Williamson's view, any knowledge worth the name evinces some sort of cognitive achievement. Obviously, fans of the 'no cognitive achievement' view of self-knowledge disagree. But the disagreement here might seem not so much substantive as terminological. In one sense, Williamson has nothing to say to this species of luminist; in another, they have already conceded his central point. To the extent that luminists want to conceive of self-knowledge as an instance of knowledge more generally, the anti-luminosity argument puts pressure on them to abandon luminosity. To the extent that luminists want to preserve luminosity, they are under pressure to accept a different picture of what knowledge itself is.

---

<sup>47</sup> Of course it might turn out that these accounts, initially motivated by the appeal of luminosity, then turn out to have other theoretical virtues (elegance, fecundity, etc.) and thus support the original datum of luminosity. But then the luminist would need to do more than simply point out that various philosophers endorse a view of the phenomenal as constitutively connected to the doxastic; he would have to show why these accounts recommend themselves independently of luminosity considerations. Thanks to Jane Friedman for raising this point.

## Acknowledgments

Thanks to Jane Friedman, Jeremy Goodman, John Hawthorne, Clayton Littlejohn, Daniel Rothschild and Tim Williamson for very helpful discussion of these issues. Special thanks to an anonymous reviewer from *Philosophy and Phenomenological Research* for detailed and insightful comments on earlier drafts, and to the audience at the New Insights in Religious Epistemology Safety Workshop, Oxford.

## Works Cited

- Berker, S. (2008). "Luminosity Regained". *Philosophers' Imprint*, 8(2): 1-22.
- Bilgrami, A. (2006). *Self-Knowledge and Resentment*. Cambridge, MA.: Harvard University Press.
- Blackson, T. (2007). "On Williamson's Argument for (I) in His Anti-Luminosity Argument". *Philosophy and Phenomenological Research*, 74(2): 397 – 405.
- Brueckner, A. and Fiocco, M. O. (2002). "Williamson's Anti-Luminosity Argument". *Philosophical Studies*, 110: 285-293.
- Chalmers, D. (2003). "The Content and Epistemology of Phenomenal Belief", in Smith, Q. and Jolic, A. (eds.), *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press, 220 – 271.
- Cohen, S. (2010). "Luminosity, Reliability, and the Sorites". *Philosophy and Phenomenological Research*, 81(3): 718 – 30.
- Coliva, A. (2009). "Self-knowledge and Commitments". *Synthese*, 171(3): 365 – 75.
- Coliva, A., (ed.) (2012). *Self and Self-Knowledge*. Oxford: Oxford University Press.
- Comesaña, J. (2005). "Unsafe Knowledge". *Synthese*, 146: 395-404.
- Conee, E. (2005). "The Comforts of Home". *Philosophy and Phenomenological Research*, 70(2): 444-451.
- Conee, E. and Feldman, R. (1998). "The Generality Problem for Reliabilism". *Philosophical Studies*, 89(1): 1-29.
- Hawthorne, J. (2005). "Knowledge and Evidence". *Philosophy and Phenomenological Research*, 70(2): 452-458.
- Heal, J. (2001). "On first-person authority". *Proceedings of the Aristotelian Society CII*: 1-19.

- Leitgeb, H. (2002). "Review of Timothy Williamson, *Knowledge and its Limits*". *Grazer Philosophische Studien*, 65(1): 195-205.
- Manley, D. (2007). "Safety, Content, Apriority, Self-Knowledge". *The Journal of Philosophy*, 104(8): 403-423.
- Neta, R. and Rohrbaugh, G. (2004). "Luminosity and the Safety of Knowledge". *Pacific Philosophical Quarterly*, 85: 396-406.
- Ramachandran, M. (2009). "Anti-Luminosity: Four Unsuccessful Strategies". *Australasian Journal of Philosophy*, 87(4): 659-673.
- Shoemaker, S. (1986). *The First Person Perspective and Other Essays*. Oxford: Oxford University Press.
- Vogel, J. (2010). "Luminosity and Indiscriminability". *Philosophical Perspectives*, 24(1): 547-572.
- Weatherson, B. (2004). "Luminous Margins". *Australasian Journal of Philosophy*, 82: 373-383.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, T. (2005). "Replies to Commentators". *Philosophy and Phenomenological Research*, 70(2): 468-491.
- Wong, W. (2008). "What Williamson's Anti-Luminosity Argument Really Is". *Pacific Philosophical Quarterly*, 89: 536-543.
- Wright, C. (1989). "Wittgenstein's later philosophy of mind: Sensation, privacy and intentions". *Journal of Philosophy*, 86(11): 622-34.
- Wright, C. (1998). "Self-knowledge: the Wittgensteinian legacy", in Wright, C., Smith, B. and Macdonald, C. (eds.), *Knowing Our Own Minds*. Oxford: Clarendon Press, 13-45.
- Zardini, E. (forthcoming). "Luminosity and determinacy". *Philosophical Studies*.

Essay Two  
What's in a Norm?

## I. A dilemma for normative philosophy

Cartesian epistemology is characterised by a dialectical relationship between doubt and certainty: radical doubt about the outer world is quelled by the epistemic security of the inner. In viewing the phenomenal realm as a safe harbour from doubt, a place of privileged epistemic access, philosophy remains very much under the Cartesian spell.<sup>48</sup> But recently the spell seems to be breaking. Arguments from empirical psychology (Gopnik 1983; Carruthers 2011; Schwitzgebel 2008, 2011) and armchair epistemology (Williamson 2000; Essay One) as well as simple introspective considerations together motivate a view on which the phenomenal is not always open to view.

The rejection of the Cartesian orthodoxy has far-reaching philosophical consequences. In particular, I want to suggest, it presents normative philosophy with a challenge. For the putative transparency of certain phenomenal conditions has been thought by many to make them uniquely suited to feature in epistemic and ethical norms. It is often said that such norms should capture our intuitive judgments about agential culpability: whether agents are (ethically or epistemically) *blameworthy* for their actions. Norms that are tied to agents' phenomenal states – internalist epistemic norms and subjectivist ethical norms – are often thought to have just this feature. Since agents are always in a position to know whether they are in certain phenomenal states (the Cartesian orthodoxy goes), it is impossible to blamelessly violate internalist and subjectivist norms. Thus norms that are tied to phenomenal states seem poised to vindicate an intimacy between the *normativity perspective*, from which it matters whether agents conform to norms, and the *culpability perspective*, from which it matters whether agents are blameworthy. If we reject the Cartesian orthodoxy, however,

---

<sup>48</sup> See Schwitzgebel 2011 for a discussion of philosophers who, despite their scepticism about folk psychology, hold onto a notion of privileged first-person access. See also various essays in Coliva 2012 for defences of a Cartesian view of the phenomenal.

then it turns out that even internalist and subjectivist norms cannot vindicate an intimacy between the two perspectives. For if there are no conditions such that we are always in a position to know whether we are in them, then it is possible to unknowingly, and thus blamelessly, violate any norm.<sup>49</sup>

Ethicists and epistemologists are thus faced with a dilemma. Either we cleave to Cartesianism, or we accept that the normativity and culpability perspectives can never be perfectly reconciled. The former most likely involves embracing a kind of metanormative anti-realism that will be attractive to some but unpalatable to many. The latter involves accepting a more tragic conception of our ethical and epistemic lives, according to which *luck* plays an ineradicable role in the normative realm.

This paper proceeds as follows. In section 2, I define and briefly motivate Anti-Cartesianism, and show that it implies Normative Non-Transparency, the thesis that there are no norms whose triggering conditions are such that we are always in a position to know whether they obtain. In section 3, I explore why Normative Non-Transparency is worrying. First I discuss a tempting but, I think, ultimately unsatisfying diagnosis (3.1): namely, that Normative Non-Transparency implies that there are no *followable* or *action-guiding* norms. I then turn to the diagnosis I favour (3.2): Normative Non-Transparency frustrates the widespread desire to eliminate the possibility of blameless norm-violation and thereby bring the culpability and normativity perspectives into alignment. In section 4, I discuss what this means, respectively, for the debate between ethical subjectivists and objectivists (4.1), and epistemic internalists and externalists (4.2). While Normative Non-Transparency should be particularly worrying for subjectivists and internalists – implying that the norms they favour are not in principle

---

<sup>49</sup> See Suikkanen 2008 and Wedgwood 2010 for a discussion of a similar claim: namely, that if Williamson's anti-luminosity argument is sound, then there exist unknowable moral truths and Dummetian realism about ethics is therefore false. Also see Henriksen 1995 and Dougherty forthcoming for a discussion of the related topic of ethical vagueness.

different from those favoured by their objectivist and externalistic rivals – I explain why many objectivists and externalists should be concerned, too. I conclude section 4 by evaluating the prospects for finding an internalist- and subjectivist-friendly substitute for normative transparency (4.3). Finally in section 5, I discuss what it might mean to grasp either of the twin horns of the dilemma: cleaving to Cartesianism by endorsing a form of metanormative anti-realism (5.1), or embracing a more tragic conception of epistemology and ethics, according to which *normative luck* plays an ineradicable role in our cognitive and moral lives (5.2).

## 2. Defining Anti-Cartesianism

Following Timothy Williamson (2000, ch. 4), let us call a condition *C* *luminous* just in case whenever one is in *C* one is in a position to know one is in *C*; *absent-luminous* just in case whenever one is not in *C* one is in a position to know one is not in *C*; and *transparent* just in case *C* is both luminous and absent-luminous. Thus a condition *C* is transparent just in case one is always in a position to know whether one is in *C*.<sup>50</sup> We can now more precisely state the Anti-Cartesian thesis:

ANTI-CARTESIANISM: There are no conditions that are transparent for creatures like us<sup>51</sup>

---

<sup>50</sup> There is a trivial sense in which paradigmatically Cartesian conditions are non-transparent, since it is possible to be or fail to be in many mental state conditions while being asleep, in which case one is not in a position to know whether one is in those conditions. Defenders of Cartesianism will naturally want to make exceptions for such trivial failures of transparency; Anti-Cartesianism is the claim that there are failures of transparency beyond these trivial ones even for paradigmatically Cartesian conditions.

<sup>51</sup> Are there exceptions? Williamson (2000, 107) points out that conditions that we are necessarily not in – impossible conditions – are vacuously luminous. Take the condition of *its being the case that 2 plus 2 equals 5*. Since that condition never obtains, it is trivially true that whenever it obtains we are in a position to know it obtains. But even such conditions won't be absent-luminous, and so won't be transparent; plausibly, we can fail to be in a position to know that 2 plus 2 doesn't equal 5 – say, if we have strong evidence that we're on a drug that prevents one from doing simple sums. What about a condition that obtains in every case, i.e. the necessary condition? Such a condition is vacuously absent-luminous, since it is trivially true that whenever it doesn't obtain we are in a position to know it doesn't obtain. But the necessary condition won't be luminous since it could be

Anti-Cartesianism is not merely the claim that for any condition we can sometimes fail to know whether we are in it. For even Cartesians will agree that we can sometimes fail to know whether we are in a given condition because of inattentiveness or negligence. Anti-Cartesianism is the stronger claim that for any condition we can sometimes fail to be even in a *position* to know whether we are in it. For no condition will assiduous attention suffice to guarantee knowledge of whether it obtains.

Non-transparency is an unsurprising feature of most conditions. Take the condition of my car being parked on Main Street, CAR. It is possible for the condition to obtain without my being in a position to know that it does.<sup>52</sup> For I might have weak evidence (only a vague memory of having parked the car on Main Street), or misleading evidence (a false report of its having been stolen). Like countless other conditions – *that it's raining in Paris, that Peter is coming to the party, that the Prime Minister is in London* – one can fail to know whether CAR obtains simply because the world and one's position in it preclude such knowledge. The radicalism of Anti-Cartesianism lies in its insistence that what is true of CAR is true of practically all conditions. This includes even paradigmatically Cartesian conditions like *feeling cold, possessing evidence that p, believing that p* and *its seeming to one that p*.

Why endorse Anti-Cartesianism? Alison Gopnik (1983), Peter Carruthers (2011) and Eric Schwitzgebel (2008, 2011) all argue that the empirical psychology data demand a rejection of Cartesianism. Both Gopnik and Carruthers deploy empirical considerations to support Ryle's thesis that self-knowledge is not

---

presented under a guise that prevents us from seeing its tautologous nature, e.g. an *a posteriori* necessity like "water is H<sub>2</sub>O". In section 5.1 I discuss whether conditions that are constitutively connected to doxastic conditions might be a plausible exception to Anti-Cartesianism.

<sup>52</sup> Without collecting more evidence, that is – I could always go to Main Street and check whether my car is there.

different in kind from knowledge of the mental states of others. Thus they argue that we can sometimes be mistaken, even in favourable conditions of sustained reflection, about our own mental states. According to Schwitzgebel, the empirical data suggest that such Cartesian failures are neither marginal nor unusual; our grasp on our mental lives – our visual imagery, our emotions, our tactile experience – is systematically *worse* than our grasp on the external world. So not only are we wrong to think we have privileged first-personal epistemic access, we moreover cannot retreat to the weaker thesis that we have at least greater epistemic access to our minds than to external states of the world.

As I argued in Essay One, Williamson's anti-luminosity argument (2000, ch. 4) gives us strong reason to think that there are no non-trivial luminous conditions – and thus, *a fortiori*, no transparent conditions. I won't rehearse my defence of Williamson's argument here. But I think it's worthwhile noting that even simpler armchair arguments in favour of Anti-Cartesianism are available. Suppose that, under the sway of psychological priming or hypnosis, one ended up systematically mistaking feelings of lukewarmth for feelings of cold; one would thereby lose the ability to know, at least in cases near the lukewarm/cold border, whether one was cold. Or suppose that one were presented with an extremely powerful philosophical argument to the effect that one's judgments about one's own phenomenology were systematically mistaken; one might thereby lose one's ability to form justified beliefs about, and thus know, one's own phenomenology. Inasmuch as one thinks these are genuine possibilities, one has reason to embrace Anti-Cartesianism.

Perhaps the most powerful reason to embrace Anti-Cartesianism is, somewhat ironically, introspective. I sometimes find myself uncertain, even after careful consideration, about my own phenomenology: whether I'm angry or merely annoyed, whether I'm desirous or indifferent, whether I believe or am agnostic.

Of course, the uncertainty at issue here is not an uncertainty about whether my phenomenology is *thus*: I'm always in a position to know that I'm feeling just *this*, the way I'm feeling. Instead, the uncertainty lies in the categorisation of my phenomenal experience under the appropriate concepts: *anger*, *annoyance*, *desire*, *indifference*, *belief*, *agnosticism*. My own introspective feelings of uncertainty deepen when we move to conditions of particular philosophical interest, such as *my having a credence  $x$  in  $p$*  or  *$p$ 's having probability  $x$  on my evidence*. For these conditions, I very often feel that no amount of assiduous introspection will reveal whether they obtain. Perhaps not everyone finds their own phenomenology so recalcitrant to his or her epistemic grasp. But to the extent that this kind of experience strikes one as familiar, one again has reason to embrace Anti-Cartesianism.

Of course, embracing Anti-Cartesianism doesn't mean accepting the sceptical verdict that one is never in a position to know whether one is in various mental state conditions – any more than it means embracing the sceptical verdict that one can never know the location of one's car. One very often is in a position to know whether one's car is on Main Street, and one might very well often be in a position to know whether one is in various mental state conditions. We might say that conditions like CAR and feeling cold are *contextually transparent*, meaning that at certain contexts one can know whether they obtain. But contextual transparency is not the same as transparency *simpliciter*. According to Anti-Cartesianism, no conditions are transparent *simpliciter*.

### 3. Normative Non-Transparency

A *norm* is a universal generalisation about how one ought to conduct one's practical or doxastic affairs, involving a normative state N and a *triggering*

condition C, of the form “N if/and/only if C”. Here are some toy examples, drawing on some of the major ethical and epistemic theories:

ACT CONSEQUENTIALIST NORM (ACN): One ought to  $\phi$  iff  $\phi$ -ing would maximise the good<sup>53</sup>

KANTIAN NORM (KN): One is permitted to  $\phi$  iff  $\phi$ -ing is in accordance with a maxim which one could at the same time as  $\phi$ -ing will that it become a universal law

EVIDENTIALIST NORM (EN): One is justified in coming to believe  $p$  iff  $p$  has a probability of at least 0.7 on one's evidence

RELIABILIST NORM (RN): One is justified in coming to believe  $p$  iff if one were to believe  $p$ , that belief would be produced by a reliable mechanism

In the case of (ACN), the normative state is *that one ought to  $\phi$* , and the triggering condition is *that  $\phi$ -ing would maximise the good*. In the case of (KN), the normative state is *that one is permitted to  $\phi$*  and the triggering condition is *that  $\phi$ -ing is in accordance with a maxim which one could at the same time as  $\phi$ -ing will that it become a universal law*. In the case of both (EN) and (RN), the normative state is *that one is justified in coming to believe  $p$* , and the triggering conditions are, respectively, *that  $p$  has a probability of at least 0.7 on one's*

---

<sup>53</sup> As with all my example norms, this will need some refinement. First, I ignore the possibility of tied best outcomes. Second, I don't control for the worry that in some cases if one were to perform some action that one ought to do one would do it in such a way that would not maximise the good. For example, it might be the case that John ought to donate most of his money to charity because doing so would maximise the good, but also the case that if John were to do so he would become suicidally depressed, thereby seriously harming himself and those who care about him (and thus, let's suppose, lowering the overall good). In that case (ACN) wouldn't give us the right (act consequentialist) verdict. Similar refinements would be required for all my example norms.

evidence and that one's belief that  $p$  would be produced by a reliable mechanism.<sup>54</sup>

(ACN) is an *obligatory* norm while (KN), (EN) and (RN) are *permissive* norms. Whether a norm is permissive or obligatory depends on the character of the normative state it describes. States like *ought  $\phi$*  and *rationally required to  $\phi$*  are obligatory states, and norms that involve them are in turn obligatory norms. Meanwhile, *permissible to  $\phi$*  and *epistemically justified in  $\phi$ -ing* are permissive states, and thus contribute to permissive norms. The conditions for satisfying permissive and obligatory norms differ. Suppose S is considering whether to perform some action  $\psi$ ,<sup>55</sup> and let us suppose that  $\psi$ -ing would in fact maximise the good. For S to satisfy (ACN), she must  $\psi$ ; a failure to  $\psi$  constitutes a violation of (ACN).<sup>56</sup> Now imagine that  $\psi$ -ing would in fact not maximise the good. For S then to satisfy (ACN), she must not  $\psi$ ;  $\psi$ -ing would constitute a violation of (ACN). Compare the permissive norm (EN). Suppose that S is deciding whether to believe some proposition  $q$ . In fact,  $q$  is 0.8 probable on S's evidence, so S is permitted to believe it. But S need not believe  $q$  to satisfy (EN); she can equally withhold or suspend judgment about  $q$ . A violation of (EN) only occurs when S believes a proposition that is less than 0.7 probable on her evidence.

---

<sup>54</sup> While moral norms are best understood as 'doing' norms, it's plausible that epistemic norms – particularly justification norms – are best understood as 'being' norms, i.e. norms that say what states one is or is not permitted/required to find oneself in. For the sake of simplicity I have written both (EN) and (RN) as doing norms, i.e. norms about what belief transitions are permissible.

<sup>55</sup> I'm assuming throughout that  $\phi$  in my toy norms quantifies over basic actions that one knows how to perform and also knows whether one is performing, and that  $\psi$  is an arbitrary action of this kind. Of course it follows from Anti-Cartesianism that one is not always in a position to know whether one is performing a given action-type. I ignore this complication, though it is relevant to showing why non-conditional norms – e.g. "One must never murder" (see n. 57) – are non-transparent (viz. since one does not always know whether one is committing murder, one does not always know whether one is violating that norm). I'm also making the analogous assumptions about my belief-norms.

<sup>56</sup> I mean for the norm "Ought  $\phi$  iff C" to be read as "If C, one ought  $\phi$ ; and if not-C, one ought not  $\phi$ ". It's an interesting question in semantics – one I won't get into – how one secures that reading. I mean permissions to be read with a narrow scope, i.e. "Permitted to  $\phi$  iff C" as "Permitted ( $\phi$ ) iff C".

A final important distinction is between norms that are biconditional and those that are merely sufficient or merely necessary.<sup>57</sup> The four norms above – (ACN), (KN), (EN) and (RN) – are all biconditional, i.e. of the form “N iff C”. But we can also have merely sufficient norms of the form “N if C”, and merely necessary norms of the form “N only if C”. These non-biconditional norms will also come in both obligatory and permissive flavours. Together these two distinctions give rise to a total of six possible forms of norm. The following table summarises these forms and indicates what constitutes a violation for each:

Biconditional/ Sufficient/Necessary	Obligatory/ Permissive	Example	Violation(s)
Biconditional	Obligatory	“S ought $\phi$ iff C”	(C & $\neg \psi$ ), ( $\neg$ C & $\psi$ )
Sufficient	Obligatory	“S ought $\phi$ if C”	(C & $\neg \psi$ )
Necessary	Obligatory	“S ought $\phi$ only if C”	( $\neg$ C & $\psi$ )
Biconditional	Permissive	“S is permitted to $\phi$ iff C”	( $\neg$ C & $\psi$ )
Sufficient	Permissive	“S is permitted to $\phi$ if C”	-
Necessary	Permissive	“S is permitted to $\phi$ only if C”	( $\neg$ C & $\psi$ )

Fig. 1 *Types of Norms*

For the sake of simplicity I will speak for the most part as if all norms are biconditional in form. But what I have to say applies to the other types of norm as well, with the exception of sufficient permissive norms, which are inviolable.

Finally, let us call a norm *transparent* just in case it features a transparent triggering condition. A corollary of Anti-Cartesianism is then:

---

<sup>57</sup> We might also have *non-conditional* norms, e.g. “One must never lie” and “One should only believe truths”. Plausibly all such unconditional norms can be re-written as conditional norms, e.g. “One is permitted to  $\phi$  iff  $\phi$ -ing isn’t an instance of lying” and “One is justified in coming to believe  $p$  only if  $p$ ”. Though I focus mainly on conditional norms, what I say applies equally to non-conditional norms (see n. 55).

NORMATIVE NON-TRANSPARENCY: there are no transparent norms

It's important to note that adding a knowledge constraint to a norm's triggering condition won't generate counterexamples to Normative Non-Transparency. For example, if I amend "One ought to  $\phi$  iff C" to "One ought to  $\phi$  iff one *knows* that C obtains", I won't have thereby created a transparent norm. For assuming that knowledge itself is non-transparent, the condition *that I know that C obtains* will be such that I can fail to know whether it obtains. Similar things can be said of adding other epistemic constraints to triggering conditions, for example that one *believes* that C obtains, or *justifiably believes* that C obtains, or that *it's probable on one's evidence* that C obtains. If Anti-Cartesianism is true, then internalising or subjectivising triggering conditions – retreating from states of the external world to related phenomenal states – won't secure transparent norms. Of course, since many conditions are contextually transparent, many norms will also be contextually transparent. But if Anti-Cartesianism is true, there are no norms that are transparent *simpliciter*.

### 3. Worrying about normative non-transparency

What does it matter if there turn out to be no transparent norms? Transparency is often taken to be a desideratum of norms. When made explicit, the importance of this desideratum is usually defended on the basis of considerations of *followability* and *action-guidance*. I find this explanation unconvincing. In section 3.1 I explain why. In section 3.2 I turn to my preferred explanation of why transparency is an intuitive desideratum of norms: namely, that transparent norms are inviolable through bad luck, which in turn means they secure an alignment of the culpability and normativity perspectives.

### 3.1 Guidance and Followability

In normative theory, one often hears the demand that norms be *action-guiding*, *followable* or *useful*.<sup>58</sup> This demand often issues from an insistence on the 'first personal' role of norms, as things that not only provide us with third-personal metrics for assessing others and states-of-affairs, but also as things that tell us how to think and act – that guide and advise us. We might think that it is precisely this demand that non-transparent norms fail to meet. After all, if we can't always know whether a norm's triggering condition obtains, how can we follow, or be guided by, or use such a norm? Thus Jim Pryor writes:<sup>59</sup>

If a belief-guiding recipe [of the form 'In circumstances C, believe *p*'] is to be *usable* in deciding what to believe, then the circumstances C it refers to must be circumstances such that the subject *can tell* whether they obtain, when he's following the recipe (2001, 116).

Similarly, Frank Jackson writes:

[T]he fact that a course of action would have the best results is *not in itself a guide to action*, for a guide to action must in some appropriate sense be present to the agent's mind. We need, if you like, a story from the inside of an agent...and having the best consequences is a story from the outside (1991, 466-7, italics added).

James Hudson explains the value of subjectivist moral theories in a similar vein:

An objective theory lays down conditions for right action which an agent may often be unable to use in determining her own behaviour. In contrast, the conditions for right action laid down by a subjective theory guarantee the agent's ability to *use them to guide her actions* (1989, 221).

---

<sup>58</sup> One finds this desideratum invoked and discussed across normative philosophy, especially in response to consequentialist and virtue theories in ethics and externalist theories in epistemology. For a few examples see Williams 1981a; Nagel 1989, 139; Bennett 1998; Shapiro 1998; Hursthouse 1999; Cruz and Pollock 1999; Gibbard 2005; Farrelly 2007; Thomson 2008 – as well as those texts quoted below.

<sup>59</sup> Voicing a view he doesn't himself endorse.

And here is Shelly Kagan, characterising how the demand for followability and guidance is deployed as a reason to be critical of act consequentialism:

Perhaps the most common objection to consequentialism is this: it is impossible to know the future. This means that you will never be absolutely certain as to what all the consequences of your act will be...In fact lacking a crystal ball, how could you possibly tell what all the effects of your act will be? So how can we tell which act will lead to the best results overall – counting all the results? This seems to mean that consequentialism *will be unusable as a moral guide to action*...[I]f it is impossible to tell whether any act is morally right or wrong, how can consequentialism possibly be a correct moral theory? (1998, 64, italics added).

If followability and action-guidance are adequacy conditions on norms, then we might think that Normative Non-Transparency is problematic because it implies that no norms are adequate.

We can unpack this diagnosis further. A norm is *operationalisable* just in case, if one knows the norm,<sup>60</sup> one is always able to use the norm as a premise in a competent deduction, together with one's knowledge of whether the norm's triggering condition obtains, to come to know whether one is acting in conformity with the norm.<sup>61</sup> Non-transparent norms are non-operationalisable.<sup>62</sup> This is because for any non-transparent norm, it is possible to fail to be in a position to know whether the norm's triggering condition obtains, and thus fail to be in a position to be able to perform the relevant deduction.

Take for example the obligatory norm "One ought  $\phi$  iff C". Suppose for some action  $\psi$ , C obtains but I'm not in a position to know it does. If I  $\psi$ , I will conform to the norm but won't be in a position to deduce that I'm so conforming; if I don't  $\psi$ , I will violate the norm, but again won't be in a position to deduce that I'm so

---

<sup>60</sup> In the 'concealed question' sense of knowing a norm, i.e. knowing what a norm says. A norm need not be true to be in this sense *known*.

<sup>61</sup> I'm assuming that one can come to know the conclusion of a competent deduction only if one knows all the premises.

<sup>62</sup> One exception: norms that are unknowable are trivially operationalisable.

violating it.<sup>63</sup> So as I decide whether to  $\psi$ , I cannot use my knowledge of the norm to deduce whether  $\psi$ -ing would satisfy it, since I do not know that the triggering condition obtains.<sup>64</sup> Take now the permissive norm “One is permitted to  $\phi$  iff C”. Suppose that for some action  $\psi$ , C doesn't obtain but I'm not in a position to know this. If I  $\psi$  I will be in violation of the norm, but won't be in a position to deduce this. Only when we are always in a position to know whether a norm's triggering condition obtains are we always in a position to deduce whether we are in conformity with the norm: operationalisability requires transparency.<sup>65</sup>

So now we are in a position to elaborate the ‘followability/guidance’ diagnosis of the importance of normative transparency. If there are no transparent norms, then there are no operationalisable norms. But if there are no operationalisable norms, then there are no norms that are followable or action-guiding. Since followability and action-guidance are adequacy conditions on norms, Anti-Cartesianism implies that no norms are adequate.

Despite the ubiquity and *prima facie* appeal of this line of thought, I find it unhelpful. My first concern is that, at least as we ordinarily deploy the concepts, *followability* and *guidance* don't seem to require operationalisability. Consider the following norm:

SEDER NORM (SN): When setting the table for Passover, one ought  
to set as many places as there will be Seder guests plus one

---

<sup>63</sup> This assumes that  $\psi$ -ing itself doesn't improve S's epistemic status vis-à-vis C, and again (see n. 55) that one is in a position to know whether one is  $\psi$ -ing.

<sup>64</sup> There is of course a sense in which, by knowing the norm, I *always* know what I ought to do to satisfy it – the issue here is knowing whether I am conforming to the norm.

<sup>65</sup> There might be some unusual cases in which an agent is always in a position to know whether she is conforming to a norm despite the norm's not being operationalisable for her. Take for example the norm “One ought to  $\phi$  iff  $\phi$ -ing will maximise the good”, and imagine a creature S who is built in such a way such that she  $\psi$ 's iff  $\psi$ -ing will maximise the good (regardless of whether she *knows* it will maximise the good), and she moreover knows that this is true. Such a (strange) creature will be able always to know whether she is conforming to the norm, although the norm is non-operationalisable for her.

(SN) seems an uncontroversial candidate for a norm that is followable and guiding. However, it is non-transparent and thus non-operationalisable. For there are occasions when one is simply not in a position to know how many people will come to one's Seder: a guest might unexpectedly fail to show up, or a surprise guest might arrive. In such cases, one is unable to know how many places to set through a deduction that uses (SN) as one of its premises. But it seems strange to conclude that (SN) is non-guiding or unfollowable.<sup>66</sup> In many situations, after all, one does know how many Seder guests one will have, and one is able to use this knowledge to deduce how many places to set. Here at least mere contextual operationalisability seems to suffice for followability and guidance.<sup>67</sup>

But we can say something even stronger. Imagine a norm that is not only non-operationalisable but moreover not even contextually operationalisable: a norm that one can *never* use in a competent deduction to come to know whether one is conforming to it. Call such a norm *opaque*.<sup>68</sup> For example, take a norm that is meant to apply to all ticket-holders in a fair lottery:

FERRARI NORM (FN): One ought to buy a Ferrari today iff one's ticket will be drawn as the winner tomorrow

Plausibly, one is never in a position to deduce whether one is conforming to (FN), since one is never in a position to know whether one's ticket will be a winner. Nonetheless, one is generally in a position to know that one's ticket will *probably* lose – and thus that one can *probably* conform to (FN) by abstaining from buying

---

<sup>66</sup> For a similar argument, see Wedgwood 2002, 365-6.

<sup>67</sup> We might hope to modify the adequacy condition on norms to demand not that they be operationalisable *simpliciter*, but rather that they be contextually operationalisable *in contexts that are salient when one asks questions like* “What's the correct moral theory?” or “How ought I to live?” Though see my next claim, viz. that even norms that are not contextually operationalisable can be followable/guiding.

<sup>68</sup> See Lenman 2000 and Kagan 1998 for an argument that the act consequentialist norm is opaque.

a Ferrari today. Despite its opacity, (FN) seems intuitively guiding and followable, for it can be meaningfully incorporated into one's practical reasoning.

What sort of incorporation into one's practical reasoning need be possible for a norm to count as guiding and followable? I don't propose to offer an account here; indeed, doing so is notoriously thorny.<sup>69</sup> But plausibly such incorporation need not involve coming to know whether one is violating the norm through a deduction from known premises. If we like we can of course reserve the honorifics 'followable' and 'action-guiding' for norms that are operationalisable. But then we won't, I think, be drawing on our commonsensical grasp of those concepts.

The second worry for the followability/action-guidance diagnosis of the importance of normative transparency is that it presupposes a picture of normative conduct as a *rule-governed* affair. That is, it assumes that conducting our practical and doxastic lives is a matter, when all goes well, of actively applying general rules to specific cases. But one might instead think that the life well lived involves being well-disposed to conform to certain norms, generally oriented towards certain *telé*, or being sensitive to the right epistemic and moral reasons. On such an account, proper human conduct need not involve any active cognising of general rules – indeed, it might even be at odds with such rule-application.<sup>70</sup> If one favoured such a view, then the (putative) revelation that no norms are followable or guiding would not be much reason for concern.<sup>71</sup>

---

<sup>69</sup> See, e.g. Wittgenstein 1953, §185-243; Kripke 1982; Boghossian 1989.

<sup>70</sup> Consider, for example, Bernard Williams' case of a man who is faced with a choice of saving the life of only one person of a group, and chooses to save his wife: "It might have been hoped by some (for instance, his wife) that his motivating thought, fully spelled out, would be the thought that it was his wife, not that it was his wife and that in situations of this kind it is permissible to save one's wife" (1981b,18). (Tim Williamson observed in conversation that the man's thought, fully spelled out, should be not *that's my wife* but *that's her*.)

<sup>71</sup> That said, if Anti-Cartesianism is correct, it will have important implications for how sensitive it is possible for creatures like us to be to the right reasons. Such sensitivity requires some sort of reliable connection between how agents behave and the facts about what reasons they have. One way of securing such a connection is cognitive: agents are aware of the relevant facts and this awareness triggers the relevant dispositions to behave appropriately in response to those facts. But there are other ways that such a reliable connection might exist. Young children and animals, for example, can demonstrate empathetic sensitivity to the suffering of others without, we might think,

In the next section I offer an alternative diagnosis of the importance of normative transparency, one that does not presuppose a rule-governed conception of the normative. According to my diagnosis, normative transparency has the potential to control the role that luck plays in the satisfaction of norms, thereby bringing the normativity and culpability perspectives into better alignment.

### 3.2 Normative luck and the culpability perspective

When making culpability judgments – that is, judgments about whether agents are blameworthy for their actions – we typically attend to those features of the situation that are within the agent's sphere of control. For what lies outside the agent's control is a matter of luck,<sup>72</sup> and intuitively the culpability facts do not turn on matters of mere luck.

This broad contention requires refinement. For in many cases our culpability judgments *do* turn on matters of luck. For example, many find it intuitive to judge people blameworthy for behaviour or behavioural dispositions that were not the result of their free will, but rather the result of natural events and laws.<sup>73</sup> And many also find it intuitive to judge agents blameworthy for behaviour that results

---

having the thought that *someone is suffering*. How such sensitivity can be secured turns on empirical questions about the agent and the kind of reason-facts at issue. In some cases, it seems plausible that sensitivity can only be secured cognitively – suppose, for example, that morality demanded that we be sensitive to facts about the atmospheric conditions of distant planets. Where sensitivity can only be secured cognitively, Anti-Cartesianism will threaten its possibility.

<sup>72</sup> Duncan Pritchard (2004) objects that, if luck is to be understood this way – as being whatever is outside an agent's control – then we will have to conclude that states-of-affairs like the motion of the planets are lucky. (Cf. Williamson 2007, 23 for similar worries about the ambiguity of 'lucky' and the resulting difficulties of thinking of epistemic internalism as the search for a 'luck-free zone'.) I don't find this is a problem myself. Our use of the terms 'luck', 'lucky' and 'unlucky' are context-dependent, but there is a clear sense in which we can say that all facts that lie outside the control of agents are in some sense 'lucky', including facts about the natural laws. In this sense, being (or failing to be) in a sceptical scenario is also a matter of luck.

<sup>73</sup> This is true of compatibilists, but also true of those who make culpability judgments in the epistemic sphere, since plausibly the case for epistemic voluntarism is quite a bit worse than the case for action voluntarism. For a discussion of epistemic voluntarism see, e.g. Heil 1984; Naylor 1985; Cook 1987; Alston 1989, Essay 5; Feldman 1988; Bennett 1990; Montmarquet 1993, ch. 5; Plantinga 1993a, ch. 2; Scott-Kakures 1994; Radcliffe 1997; Steup 1988, 1999.

from unlucky ignorance of the fundamental normative truths – say, being raised in a culture that teaches that slavery is permissible.<sup>74</sup> And most of us find it intuitive to judge people blameworthy for actions that they had the opportunity to perform only because of the particular circumstances in which they found themselves: think of the concentration camp guard who would have lived a perfectly harmless life in more mundane circumstances. To a greater or lesser extent, these forms of luck are generally felt to be compatible with the culpability perspective.<sup>75</sup>

But there is a certain kind of luck that is generally felt to be particularly at odds with the culpability perspective. This is the kind of luck discussed by Bernard Williams (1976, 1981a, 1993a) under the heading of ‘moral luck’, by Thomas Nagel (1979) as ‘resultant luck’, and by T.M. Scanlon (2008) as ‘outcome luck’. This form of luck is present when one’s ignorance of the relevant circumstantial facts prohibits one from controlling the outcome or result of one’s actions.<sup>76</sup> Suppose Kate, a generally thoughtful person, brings some soup to her friend Mary, who is down with the flu. And suppose further that Kate isn’t in a position to know that the soup contains an allergen that, rather than making Mary feel better, will make her more ill. In trying her best to help her friend, Kate inadvertently makes her friend worse off.

Intuitively, Kate is blameless for making Mary worse off, for she couldn’t have known that this would be the result of her action. Kate is a victim of bad resultant luck,<sup>77</sup> and this bad luck excuses her.<sup>78</sup> Williams writes that the very idea of

---

<sup>74</sup> On moral ignorance, and its relationship to our culpability judgments, see, e.g. Smith 1983; Zimmerman 1997; Rosen 2004; Harman 2011.

<sup>75</sup> Another, under-discussed form of normative luck: the extent to which an agent is able to take advantage of what she’s in a position to know about what she ought to do, given the other normative demands on her. For one might be in a position to know that one ought to  $\psi$  and also be in a position to know that one ought to  $\phi$ , but not be in a position to know that one ought to both  $\psi$  and  $\phi$ . Being burdened by a number of normative requirements is a form of bad luck that can stop one doing what is required of one.

<sup>76</sup> This presupposes the non-sceptical view that some actions *are* within agential control.

<sup>77</sup> When Williams, Nagel and Scanlon speak of the phenomenon of resultant luck, they have in mind both the presence and *absence* of the kind of luck we see in the Kate case. So we can contrast

resultant luck “suggest[s] an oxymoron”, because immunity from resultant luck is “built into” morality (1993a, 251-252).<sup>79</sup> From the culpability perspective, what matters is what lies within agential control; the unknowable, hazardous contribution of the external world does not come into it. In Kantian terms, what matters is the *good will*, not what unknowable circumstances might determine its results will be:

A good will is not good because of what it effects or accomplishes, because of its fitness to attain some proposed end, but only because of its volition, that is, it is good in itself...Even if, by a special disfavor of fortune or by the niggardly provision of a step motherly nature, this will should wholly lack the capacity to carry out its purpose — if with its greatest efforts it should yet achieve nothing and only the good will were left (not, of course, as a mere wish but as the summoning of all means insofar as they are in our control) — then, like a jewel, it would still shine by itself, as something that has its full worth in itself (1785/1959, 4:394).

Hume tells us that what matters from the culpability perspective is an agent's *motive*, to which an agent's normative performance is at best a defeasible guide:

‘Tis evident, that when we praise any actions, we regard only the motives that produced them, and consider the actions as signs or indications of certain principles of mind and temper. The external performance has no merit. We must look within to find the moral quality. This we cannot do directly; and therefore fix our attention on actions, as on external signs. But these actions are still considered as signs; and the ultimate object of our praise and approbation is the motive, that produced them (1739/1978 III.II.1).

---

Kate with Kate\*, who is just like Kate except that the soup she brings her friend doesn't contain an allergen. We might say that Kate\* benefits from the *absence* of bad resultant luck. Thus Kate\* is a beneficiary of resultant luck while Kate is its victim. For the most part I use the term ‘resultant’ luck in the more restricted way, i.e. simply to mean the *presence* of bad resultant luck. I explain why shortly.

<sup>78</sup> On excuses, see Austin 1956.

<sup>79</sup> Though both Williams and Nagel point out that our actual practices of blaming and punishing suggest that we are more comfortable with moral luck than we might theoretically admit. We think the negligent driver who runs over a child did something worse – indeed, is more blameworthy – than the negligent driver whose way happened to have been clear. And (Williams claims) we think that Gauguin was justified in abandoning his family because he turned out to be a great artist; had his artistic endeavours produced only mediocre results, we would judge his actions – and his blameworthiness – more harshly.

In a similar vein, we might say that what matters from the culpability perspective is whether agents *try their best*, not whether they achieve the best. Since Kate tried her best to help her friend – though she failed to achieve the best for her – she is blameless.

Crucially, resultant luck introduces a mismatch between the normativity and culpability perspectives – that is, a mismatch between whether someone has conformed to a norm, and whether someone is blameworthy. Recall the Act Consequentialist Norm, (ACN). What verdict does (ACN) pass on Kate's action? Let us suppose that the only actions available to Kate were giving Mary the (allergen-contaminated) soup or not. And let's further stipulate that making Mary worse off reduces the overall good. Thus, in giving Mary the soup, Kate violates (ACN). Thus it is possible to blamelessly violate (ACN). More generally: any norm that can be violated through bad resultant luck is a norm that can be violated blamelessly.

It will come as little surprise, especially to critics of consequentialism, that (ACN) is violable through resultant luck. Similarly, it will come as little surprise, especially to critics of externalism, that the Reliabilist Norm (RN) can be similarly violated through bad luck.<sup>80</sup> Take Henrik, who generally does his best to form only beliefs that are reliable. Acting on his disposition, he forms the belief that he has hands. Unluckily for Henrik, he has recently become a brain-in-a-vat. Since he cannot know that his perception is unreliable, Henrik cannot know that he is violating (RN) by believing he has hands. He is intuitively blameless for violating (RN).

The possibility of blameless violation is a familiar upshot of objectivist and externalist norms. What is more surprising is that, assuming Normative Non-Transparency, similar things can be said of all other norms, including those that

---

<sup>80</sup> From now on, when I speak of 'luck', I mean specifically *resultant* luck, unless otherwise noted.

we tend to think of as immune from bad luck. Take the Kantian Norm, (KN). Suppose that Sam must decide whether to lie to a suspected murderer about the whereabouts of a friend. Sam tries his best only to perform actions that are in accordance with universalisable maxims. Let's stipulate that according to (KN), Sam's lying to the would-be murderer is impermissible. However, Sam isn't in a position to know this: he has, say, read lots of conflicting Kantian scholarship on the question, and his evidence underdetermines the answer. Trying his best to conform to (KN), Sam decides to lie. In so doing, he violates (KN), but through bad luck. Intuitively, he is blameless for so doing. Or take the Evidentialist Norm, (EN). Suppose Sandra is disposed to conform to (EN). That the butler is guilty has a 0.699 probability on Sandra's evidence, and so according to (EN) she should not believe it. But for all Sandra knows, her evidence supports that proposition just above the threshold for permissible belief. Her disposition to conform to (EN) results in her believing that the butler did it. She thereby blamelessly violates (EN).

If Anti-Cartesianism is true, then all norms turn out to be susceptible to resultant luck. Any norm can be violated because one isn't in a position to know whether the relevant triggering condition obtains. In other words, any norm can be blamelessly violated. To the extent that we think that the culpability perspective is central to normativity – that ethics and epistemology are in the business of securing and explaining our judgments of agential blameworthiness and blamelessness – then we have reason to be worried by Anti-Cartesianism. For Anti-Cartesianism implies that we will be unable to construct norms that align the normativity and culpability perspectives.

My focus here has been on the phenomenon of blameless violation, which occurs whenever an agent violates a norm through bad resultant luck. But what about the possibility of conforming to norms through *good* resultant luck? Suppose Sveta

believes that her dreams are prophetic, despite having overwhelming evidence that prophecy is impossible and that she has never correctly prophesied the future. On the basis of a dream she has, Sveta forms the belief that it will snow the next day. In fact, Sveta is indeed prophetic, and it indeed does snow the next day. In forming the prophetic belief, Sveta conforms to the Reliabilist Norm, (RN), since the belief is a result of a reliable mechanism. Nonetheless, we might intuitively want to say that Sveta's compliance is a matter of good luck, since she behaved negligently in ignoring the evidence of her own unreliability.<sup>81</sup> Sveta's conformity to (RN), we might say, is *incused*; it does not render her epistemically worthy.

Sveta's negligent conformity is made possible by (RN)'s non-transparency, viz. the possibility of Sveta being in a position to justifiably but falsely believe that she's violating (RN). Were one always in a position to know whether one's belief were the result of a reliable mechanism, such a case could not arise. One might think, then, that a transparent norm could screen off all cases of negligent conformity.

But plausibly not. Recall Kate and her ailing friend Mary. Suppose Kate is in a position to know that giving Mary soup would make her feel better. But Kate is utterly indifferent to this fact; she really doesn't care that Mary is feeling poorly. Instead, she brings Mary soup because she hopes to have a chance to flirt with her brother. In bringing Mary the soup, Kate makes Mary feel better (and, we might suppose, her brother too), thereby conforming to (ACN). Intuitively though, this conformity is negligent, since Kate was indifferent to the pursuit of the good. But notice that this is not an instance of normative non-transparency, since Kate was, *ex hypothesi*, in a position to know that bringing the soup to Mary would maximise the good. Thus transparent norms, if there were any, would not

---

<sup>81</sup> This is inspired by BonJour's clairvoyant cases (1980).

eradicate cases of negligent conformity. Transparent norms only promise to eradicate cases of blameless violation.

If, as Anti-Cartesianism implies, there are no transparent norms, then we cannot eradicate the possibility of blameless violation, thereby securing an alignment between the normativity and the culpability perspectives.<sup>82</sup> One might be tempted to object that we can eliminate the phenomenon of blameless violation simply by writing into a given norm the requirement that agents be in a position to know what the norm demands of them. For example: “One ought to  $\phi$  only if *one is in a position to know* that one ought to  $\phi$ ”.<sup>83</sup> But this is simply to deny Anti-Cartesianism, for it implies that whenever it is the case that one ought to  $\psi$ , one is in a position to know one ought to  $\psi$ . Perhaps normative conditions like *ought  $\psi$*  are indeed genuine counterexamples to Anti-Cartesianism; but we require an argument, not mere stipulation, that this is so.<sup>84</sup> If Anti-Cartesianism is true, it follows that there are no norms that are immune to blameless violation.

It's worth stressing that the issue here is not the elusiveness of the correct normative theory. Our ignorance of the true norms is a topic of much interest to moral philosophers especially.<sup>85</sup> Instead at issue here is a more mundane, and more fundamental, form of normative ignorance: ignorance of the contingent, normatively relevant facts of our situation. If, as Anti-Cartesianism maintains, this kind of ignorance is a structural feature of our existences, then our traditional way of thinking about the place of luck in normative theory is under pressure. The particular focus on resultant luck in normative theorising, I suspect, is due to the

---

<sup>82</sup> Since, as I have argued, even normative transparency wouldn't be able to eradicate the possibility of negligent conformity, there is necessarily a gap between the normativity and culpability perspectives in this sense. The gap I am speaking of here is only the one caused by blameless violation.

<sup>83</sup> This was suggested to me by both Anandi Hattiangadi and Krister Bykvist.

<sup>84</sup> I offer what I take to be the best argument that we ought to deny Anti-Cartesianism for normative conditions in section 5.1.

<sup>85</sup> On moral ignorance, see Smith 1983; Zimmerman 1997; Rosen 2004; Harman 2011. On moral uncertainty, see Oddie 1995; Lockhart 2000; Sepielli 2009.

fact that the other forms of luck exist (if they exist) regardless of one's particular normative theory. By contrast, it is often thought that one can control for resultant luck in one's normative theory through the articulation of transparent norms. If Anti-Cartesianism is correct, then the attempt to control for resultant luck is futile. And so resultant luck threatens to be a pervasive feature of the normative sphere.

#### 4. Normative non-transparency, ethics and epistemology

Not all normative theorists are equally committed to the project of eradicating resultant luck from their theories. Some theorists – specifically, many ethical subjectivists and epistemic internalists – are quite strongly committed to the project, hoping to articulate norms that are not susceptible to blameless violation. As such they are particularly under threat from Anti-Cartesianism. Meanwhile ethical objectivists and epistemic externalists are more sanguine about resultant luck. And yet many of them hope to eradicate at least some instances of resultant luck from at least some of their norms; to that extent, they too are threatened by Anti-Cartesianism. In the next two sections (4.1 and 4.2) I discuss how these issues play out – first in the ethical sphere, and then in the epistemic. In section 4.3 I discuss whether internalists and subjectivists can dodge the threat from Anti-Cartesianism by shifting their focus from the importance of normative transparency to some related desideratum.

##### 4.1 Ethics: subjectivism and objectivism

*Ethical subjectivism* is the view that the moral ought supervenes entirely on conditions that are 'subjectively' available to agents. Since (the Cartesian orthodoxy goes) phenomenal states are the only transparent conditions, the moral ought must supervene on phenomenal states. *Ethical objectivism* is the denial of subjectivism; according to it, the moral ought does not supervene entirely on

conditions that are transparent to agents (whether or not there exist such transparent conditions).<sup>86</sup>

Consider again the act consequentialism norm, (ACN). That one can violate it through bad luck is often taken to be a serious objection.<sup>87</sup> In particular, it is often thought reason to prefer a 'subjectivised' version of the norm:<sup>88</sup>

SUBJECTIVE ACT CONSEQUENTIALIST NORM (SACN): S  
ought to  $\phi$  iff  $\phi$ -ing has the greatest expected goodness on S's  
evidence<sup>89</sup>

(SACN) is often taken to be superior to (ACN) because one can supposedly violate the latter but not the former through bad luck. But if there are no transparent conditions, then one isn't always in a position to know which of the available actions has the greatest *expected* goodness on one's evidence, and so (SACN), like (ACN), turns out to be non-transparent. Suppose Sanja is deciding whether to donate some money to one of two charities, Oxfam and Salvation Army. She has some non-decisive evidence about the comparable effectiveness of the two charities; on Sanja's evidence, donating to Salvation Army has *just* slightly higher expected goodness than donating to Oxfam. The slimness of the evidential margin means that Sanja isn't in a position to know that donating to Salvation Army has a higher expected goodness than donating to Oxfam. Now suppose that after trying her best to evaluate her evidence, Sanja decides to donate to Oxfam.

---

<sup>86</sup> Typically objectivists, like epistemic externalists, accept that certain mental conditions are transparent to agents; their denial of subjectivism (and the externalists' denial of internalism) is motivated not by Anti-Cartesianism, but by other considerations.

<sup>87</sup> See e.g. Kagan 1998, 64; Lenman 2000; Prichard 1932/1949; Ross 1939, 146-167; Ewing 1947, 128; Gruzalski 1981; Smith 1988; Hudson 1989; Broome 1991, 128; Griffin 1996, ch. 7; Howard-Snyder 1997; Scanlon 2001, 2008; Mason 2003; Wedgwood 2003, 204; Timmons 2012, 10-11.

<sup>88</sup> Or indeed to move away from the realm of consequentialism altogether – toward, say, deontology or virtue ethics.

<sup>89</sup> This is a very rough approximation of what a good subjectivised version of consequentialism would look like, but the needed refinements aren't important for my purposes.

In so doing she violates (SACN) – but through bad luck, and thus blamelessly. (SACN), like its objective counterpart (ACN), is susceptible to blameless violation.

We might be tempted to control for this possibility with a further round of subjectivisation: Sanja ought to do what *probably* has the greatest expected goodness on her evidence. But then we just recapitulate the worry, for, assuming Anti-Cartesianism, there will be cases in which Sanja isn't in a position to know what *probably* has the greatest expected goodness on her evidence...and so on.<sup>90</sup>

Thus the subjectivised version of the act consequentialism norm is not in principle different from its objectivist rival; it is possible to blamelessly violate both. An analogous argument can be given to show that there is nothing in principle to differentiate our Kantian Norm (KN) from a subjectivised version of it. If Anti-Cartesianism is true, then one is not always in a position to know whether an action is *probably* in accordance with a universalisable maxim, or even whether one *believes* that it is so. If so, then there will be cases in which one blamelessly violates subjectivised versions of the Kantian norm. The objection so commonly voiced by subjectivists against objectivist theories – namely, that they are susceptible to blameless violation – turns out to be a *tu quoque*.

Together, ethical subjectivism and Anti-Cartesianism yield nihilism about the moral ought. If there are no 'subjectively' available conditions, then there is no suitably subjectivist supervenience base for the moral ought. Ethical subjectivists have the choice of embracing nihilism about the moral ought, or (more likely) accepting that there is no principled distinction between the norms they favour and those favoured by their objectivist rivals – that all norms are, in this sense, objectivist.

---

<sup>90</sup> Indeed we might worry that such a norm is even less transparent than the less subjectivised one. See section 4.3 for a discussion of degrees of transparency.

This is naturally unwelcome news for subjectivists. But not only for them. For many moral theorists dwell in a halfway house between ethical subjectivism and objectivism, distinguishing between two (or more) norms, at least one of which they claim to be subjective.<sup>91</sup> Thus they might say that Kate ought (in the objective sense) not bring Mary the soup, while she ought (in the subjective sense) bring Mary the soup. In a particularly subtle example of this strategy, Derek Parfit distinguishes between various senses of 'wrong': 'wrong' in the *fact-relative sense*, 'wrong' in the *belief-relative sense*, and 'wrong' in the *evidence-relative sense* (2011, ch. 7). These different senses of 'wrong' yield different senses of 'ought', and connect up with other normative notions. According to Parfit, what one has reason to do is connected with the fact-relative sense of 'wrong', but what one is rational in doing is connected with the belief- and evidence-relative senses. Meanwhile, he claims, blameworthiness connects with the belief-relative sense of 'wrong'; one is blameworthy just in case one performs an action that would be wrong "in the ordinary sense" if one's beliefs about the morally relevant facts were true (ibid., 150). Parfit explicitly rejects the fact-relative ought as providing conditions for blameworthiness on the grounds that it does not screen off resultant luck (ibid., 156-7). But Parfit is wrong to think, if Anti-Cartesianism is true, that similar objections cannot be made against the belief-relative sense of ought. If one is not always in a position to know what one believes about the morally relevant facts, then one might very well end up blamelessly violating the belief-relative ought. One cannot force the normativity and culpability perspectives into alignment by proliferating our normative concepts. Halfway house ethicists are thus, like subjectivists, under threat from Anti-Cartesianism.

As we have seen, worries about non-transparency often attend discussions of act consequentialism. As I've just argued, if Anti-Cartesianism is true, these worries

---

<sup>91</sup> See e.g. Sidgwick 1907; Parfit 1984, 24-40; Gibbard 1990, 42-3; Zimmerman 1996, 10-20.

are live for all norms, not just consequentialist ones.<sup>92</sup> But the rather obvious non-transparency of the (objectivist) act consequentialist norm means that it has garnered a lot of attention. A common consequentialist response to the transparency worry, distinct from the move toward subjectivism, is to posit an auxiliary structure, sometimes called a *decision procedure*,<sup>93,94</sup> alongside (ACN). (ACN) is said to be then merely a 'criterion of right', a specification of what makes a given action right or wrong.<sup>95</sup> It is the decision procedure to which, on this view, agents should appeal when making decisions about what to do.<sup>96</sup> Thus Mill writes:

To inform a traveler respecting the place of his ultimate destination, is not to forbid the use of landmarks and direction-posts on the way. The proposition that happiness is the end and aim of morality, does not mean that no road ought to be laid down to that goal, or that persons going thither should not be advised to take one direction rather than another (1861/2001, 24).

To what extent does the positing of decision procedures help the proponent of act consequentialism address the challenge I have laid out? If by a 'decision procedure' we mean some norm or rule that is – unlike (ACN) itself – transparent, then there is of course no such thing, assuming Anti-Cartesianism. Of course some norms might be more transparent than others, even if no norm is perfectly transparent; some norms are contextually transparent at more (and perhaps more significant or salient) contexts than others. So perhaps the decision procedure of, say, acting in accordance with commonsense morality is more transparent than

---

<sup>92</sup> For a different argument that this is true, see Smith 1988.

<sup>93</sup> See e.g. Mill 1861, ch. 2; Moore 1903, 162–4; Sidgwick 1907, 413, 489–90; Smart 1961/1973, § 7; Bales 1971; Hare 1981, ch. 4; Parfit 1984, 24–9, 31–43; Railton 1984, 140–6, 152–3; Pettit and Brennan 1986; Hooker 2000.

<sup>94</sup> This is not the only consequentialist motivation for positing decision procedures. Act Consequentialism is said to suffer from a species of self-defeat, since encouraging people to satisfy (ACN) will not lead (it is claimed) to the maximisation of the good; Sidgwick calls this the "paradox of hedonism" (1907, 405). The introduction of decision procedures is sometimes taken to remedy this problem. My argument doesn't address this motivation.

<sup>95</sup> Analogously, Conee and Feldman (2011, 310–314) argue that their evidentialist theory of epistemic justification is not meant to furnish an evaluation of agents, since their theory implies that agents can act negligently while being justified. Instead, Conee and Feldman claim only to be formulating conditions under which a particular person time-slice's beliefs fit the available evidence in the way required for the achievement of justification.

<sup>96</sup> See Bales 1971 for an account and defence of this distinction between criterion of right and decision procedure.

the norm of maximising the good. Nonetheless, if no conditions are transparent – including *being in accordance with commonsense morality* – then the decision procedure of obeying commonsensical morality is not transparent, either. So there will be cases in which one blamelessly violates not only the criterion of right, but also the decision procedure. While positing decision procedures might have some other merits,<sup>97</sup> it is not an effective strategy for closing the normativity-culpability gap.

## 4.2 Epistemology: internalism and externalism

Like moral philosophers, epistemologists can be thought of as dividing according to their attitudes toward *epistemic* luck: the extent to which factors outside agential control play a role in determining facts about epistemic justification or epistemic rationality.<sup>98</sup> Indeed, a main motivation for epistemic internalism – the view that epistemic justification strongly supervenes on states that are ‘internal’ to agents – seems to be the intuition that justification cannot be distributed by luck.<sup>99</sup> The supervenience base for epistemic justification must be confined to conditions that are transparent to agents;<sup>100</sup> otherwise, agents could possess or lack

---

<sup>97</sup> If the problem with (ACN) is that it is effectively *opaque* – that is, that agents are never in a position to know what it demands of them (cf. Lenman 2000; Kagan 2008) – then even a relatively more transparent norm might constitute some improvement to the overall consequentialist theory. And a similar story might be told about decision procedures for other normative theories. See n. 94.

<sup>98</sup> A recent trend in epistemology uses the term ‘epistemic luck’ to pick out a phenomenon that, on all accounts, is incompatible with knowledge – i.e. the kind of epistemic luck that we find in Gettier cases (Pritchard 2004, 2005, 2012; Steglich-Petersen 2009). This is not the way I’m using the term; by ‘epistemic luck’ I mean the epistemic analogue of moral resultant luck: the kind of luck that is typically considered compatible with justification (and knowledge) by externalists, but not so by internalists.

<sup>99</sup> One might think that there are issues similar to the ones I am raising here for the internalist/externalist debate in philosophy of mind. See Boghossian 1994.

<sup>100</sup> Here and throughout I use ‘internalism’ to refer to the traditional, ‘access’ variety of internalism, which maintains both an ‘internal’ supervenience thesis and a thesis about privileged first-personal ‘access’ to the internal (Chisholm 1977, ch. 6, section 5; BonJour 1985, chs. 1-2). Access internalism contrasts with ‘mental state’ internalism, which accepts the supervenience of justification on the mental, while denying (or remaining agnostic on) the access thesis (Conee and Feldman 2001). I find it hard to understand why, once we have given up the presumption of privileged access, we should limit the epistemic justifiers to (non-factive) mental states, as mental state internalists insist. Wedgwood (2002) argues that we should accept the internalist supervenience thesis (about epistemic rationality, not necessarily epistemic justification) because

justification for reasons outside their control.<sup>101</sup> Since (the Cartesian orthodoxy goes) some mental states are transparent to agents – and nothing else is – these must constitute the supervenience base for justification. Epistemic externalists meanwhile reject the internalist thesis that justification supervenes entirely on states that are internal to agents – and with it, the view that epistemic justification cannot be a matter of luck. Internalists often find the externalist's willingness to embrace luck mystifying. Here, for example, is Laurence Bonjour:

Why...should a reason that is outside the cognitive grasp of a particular believer nonetheless be taken to confer epistemic justification on his belief?...How can the fact that a belief is reliably produced...make my acceptance of that belief rational and responsible when that fact itself is entirely unavailable to me? No doubt I am quite fortunate if my perceptual and introspective beliefs happen to be caused or produced in such a way, for then such beliefs will in fact be mostly true...indeed, such cognitive serendipity might, at least if it were thoroughgoing enough, be in the end preferable from a purely practical standpoint to anything that mere reason could be very confidently expected to deliver. But none of this has any clear bearing on the issue of epistemic justification, which has to do...with the rationality or irrationality of one's beliefs, rather than with what appears to be...*mere cognitive luck* (2003, 27, italics added).

---

complete folk psychological explanations of belief and belief revision invariably feature 'internal' states. Even if Wedgwood is right about the habits of folk psychological explanation, I'm unsure how much philosophical weight this can carry; one might well worry that folk explanations bottom out at certain non-factive mental states because it is tacitly presumed that these states are transparent. Why else do these folk explanations make epistemic behaviour, in Wedgwood's terms, "intuitively intelligible" and "unsurprising" (357)? An interesting test here might be to see how a typical folk explanation of belief revision would respond to an instance of, say, doxastic non-transparency. Wedgwood also suggests that his view of epistemic rationality can vindicate the intuition that whether one's belief is rational is a matter of control, not luck (355); this is because rationality consists in 'basic rules' that can be 'followed directly'. Again, one might suspect that folk intuitions about what rules can and cannot be 'directly followed' presuppose a commitment to transparency. For a discussion of varieties of internalism and their motivations, see Pryor 2001.

<sup>101</sup> Some epistemologists who defend internalism along these lines are: Ginet (1975, ch. 3); Chisholm (1977, ch. 1, section 5, 1982); Bonjour (1985, ch. 1); Steup (1999) and Audi (2001). For a defence of the claim that deontic considerations are what motivate internalism, see Alston 1989, essay 8; Plantinga 1993a, ch. 1; Goldman 1999. For a general discussion of the connections between justification, responsibility and blame in epistemology, see Goldman 1980, section V, 1988; Kornblith 1982; Plantinga 1988, 1993a, ch. 1; Steup 1988; Alston 1989, essays 4, 5, and 8; Moser 1989, section 1.3; Fumerton 1995, ch. 1; Haack 1997; Pryor 2001; Wedgwood 2002; Williamson 2007.

A crucial difference between internalists and externalists, then, is their willingness to countenance the role of luck in the satisfaction of epistemic norms. As Linda Zagzebski writes:

The dispute between externalists and internalists looms large mostly because of ambivalence over the place of luck in normative theory. Theorists who resist the idea that knowledge...is vulnerable to luck are pulled in the direction of internalism...Externalists are more sanguine about luck (1996, 39).

While externalists might be more sanguine about epistemic luck, many resist the extreme position that all epistemic norms are susceptible to it. For instance, many externalists maintain that there are not one but two distinct concepts of justification, one internal and one external; externalist norms provide us the conditions for correct epistemic behaviour from an objective or third-personal point-of-view, while internalist norms track the epistemic worth of agents by being luck-free. Voicing such a 'pluralist' view, Richard Foley writes:

The assumption that the conditions which make a belief rational are by definition conditions that turn a true belief into a good candidate for knowledge is needlessly limiting. It discourages the idea that there are different, equally legitimate projects for epistemologists to pursue. One project is to investigate what has to be the case in order to have knowledge. An externalist approach is well suited to this project. A distinct project, also important, is concerned with what is required to put one's own intellectual house in order (2001, 21).<sup>102</sup>

Similarly, many externalist epistemologists distinguish between justification (the externalist notion) and an internalist notion of *rationality*. Thus whether one is justified or not is not a fact to which one has privileged access; but one nonetheless is always able to know whether one is believing in a rational manner. This kind of

---

<sup>102</sup> See also Mackie 1976, 217-20; Goldman 1980, 36ff., 1986, 73ff; Alston 1989, 86ff; Sosa 1991, 240; Plantinga 1993a, 15ff.; Montmarquet 1993, ch. 6; Pollock and Cruz 1999, 140; Pryor 2001, 111ff; Toulmin 2001; Alston 2005.

pluralism can be thought of as the epistemic analogue of the ethical halfway house strategy of distinguishing between an objective and subjective ought.<sup>103</sup>

Anti-Cartesianism presents both internalists and halfway house externalists with a challenge. 'Internalist' norms turn out, if Anti-Cartesianism is true, to be in fact externalist, in the sense that their triggering conditions are sometimes external to the epistemic grasp of agents. Take the (paradigmatically externalist) norm (RN). A typical internalist complaint against it is that agents aren't always in a position to know whether their belief-forming mechanisms are reliable, and therefore aren't always in a position to know whether they are in conformity with the norm. Brain-in-a-vat Henrik, for example, violates (RN) when he believes (on the basis of his unreliable perception) that he has hands. But since he isn't in a position to know that his perception is unreliable, how then – internalists want to know – could he be *unjustified* in believing he has hands?

If Anti-Cartesianism is true, similar things can be said of paradigmatically internalist norms. Recall Sandra, who is disposed to conform to (EN). Sandra violates (EN) by forming the belief that the butler is guilty, though she is in no position to know that she is violating it; for all she knows, *that the butler did it* is sufficiently probable on her evidence to permit belief. Just as with the externalist norm (RN), one can violate the internalist norm (EN) through bad luck, and thus blamelessly.

---

<sup>103</sup> Another typical externalist strategy for controlling luck is the positing of *defeat* conditions. See e.g. Goldman 1986, 62-3, 111-2; Alston 1988, 238-9; Plantinga 1993b, 40-2, 229-37; Nozick 1981, 196. This is a strategy that particularly targets the previously discussed phenomenon of negligent conformity (but not blameless violation). According to defeatist externalism, being in a justified position to believe that one's belief that *p* is unjustified renders one's belief that *p* unjustified. Thus one can never have a justified belief through extreme negligence. For an argument that Anti-Cartesianism plus defeatism yields scepticism, see Williamson 2011. For a critical response to Williamson's argument, see Goodman 2013. For a general argument against defeatism, see Aarnio-Lasonen 2011.

If Anti-Cartesianism is true, then it turns out that no conditions are 'internal' in the way required for internalism, and so there exists no supervenience base for epistemic justification. So internalism together with Anti-Cartesianism yields nihilism about epistemic justification. Internalists thus have a choice: embrace nihilism about epistemic justification, or (more likely) accept that there is no principled distinction between the norms they favour and those favoured by their externalist rivals.

### 4.3 Substitutes for transparency

I have claimed that, if we accept Anti-Cartesianism, subjectivists and internalists face a choice between embracing nihilism about epistemic justification and accepting that their favoured norms are not in principle different from the norms favoured by their rivals. Are these the only options? A third option would be to find a desideratum, distinct from but related to transparency, that all and only internalist/subjectivist norms share. Then one could claim that the internalist/externalist and subjectivist/objectivist distinctions, though not running along the fault-line of normative transparency, are nonetheless principled.

One might argue, for example, that what matters isn't normative transparency *simpliciter*, but rather *degree* of contextual transparency. For while no norms, assuming Anti-Cartesianism, are transparent *simpliciter*, some norms will be transparent at more and perhaps more important contexts than others. Assuming we have a uniform way of individuating possible contexts, we could theoretically count the number of contexts in which any given norm is contextually transparent, and could also weight each context by its relative importance. These considerations would allow us to produce an overall measure of degree of contextual transparency for each possible norm. What to do with such a measure? We might think that it is an adequacy condition on a norm that it meet a certain

threshold degree of transparency. Or we might think that degree of transparency – like degree of simplicity, elegance or explanatoriness – constitutes just one more theoretical virtue as we go about choosing our favourite norms. In either case, the internalist and subjectivist might argue that considerations of degree of transparency tell in favour of their norms, and against rival externalist and objectivist norms.

Alternatively, the internalist or subjectivist might argue that what matters isn't normative transparency *simpliciter*, but rather that norms be transparent in *non-liminal* cases. As I argued in Essay One, Williamson's anti-luminosity argument turns on our inability to know that we are in mental state conditions when there exist nearby cases in which we are not in those conditions. If the existence of such liminal cases is the only obstacle to transparency, then we might think that internalist and subjectivist norms are superior to their externalist and objectivist rivals in being transparent in all but liminal cases. We might say that internalist and subjectivist norms are *cosy-transparent*, where a norm is cosy-transparent just in case it is transparent in non-liminal cases of its triggering condition's obtaining or not obtaining.<sup>104,105</sup>

Can internalism and subjectivism be successfully defended by appealing either to degree of transparency or to cosy-transparency? A significant worry here is that these new desiderata might not tell in favour of internalist and subjectivist norms. Take degree of transparency. Once we have embraced Anti-Cartesianism, there is no armchair assurance that internalist and subjectivist norms are more transparent than their externalist and objectivist rivals; indeed, whether they are so seems mainly an empirical question. There's no armchair guarantee that we

---

<sup>104</sup> Thanks to Harvey Lederman for this thought.

<sup>105</sup> I take the term 'cosy' from John Hawthorne (2005), who argues in response to Williamson's anti-luminosity argument that while there might be no (non-trivial) luminous conditions, there might still be cosy conditions, i.e. conditions such that whenever they obtain determinately, we are in a position to know that they obtain.

are more often in a position to know, say, whether  $p$  is probable on our evidence than we are in a position to know, say, whether my belief that  $p$  was produced by a reliable mechanism.<sup>106</sup> So it's not clear that replacing transparency with degree of transparency as our normative desideratum will tell in favour of internalist and subjectivist norms.

A similar worry arises for the appeal to cosy-transparency. It's true that Williamson's anti-luminosity argument shows only that we cannot know that we are in conditions when we are in liminal cases of being in those conditions. And it's moreover true that objectivist and externalist norms fail to be transparent for reasons beyond liminality. Take our reliabilist norm, (RN), and brain-in-a-vat Henrik. Henrik is in a very central case of (RN)'s triggering condition's not obtaining, and yet he isn't in a position to know that it doesn't obtain. Contrast our evidentialist norm, (EN). One might think that it only exhibits transparency failures when agents are in liminal cases of its triggering condition obtaining or not obtaining. Indeed, the case I've discussed – Sandra's belief that the butler is guilty – is such a case.

But recall some of the simple armchair arguments that speak in favour of Anti-Cartesianism: for example, the possibility of being psychologically primed or hypnotised. Are such possibilities live for non-liminal cases of being or not being in certain conditions? Suppose one feels very cold indeed. Is it possible for a powerful hypnotist to make one invariably mistake one's feeling very cold indeed for not feeling cold? My inclination is to say yes, though others will no doubt have the intuitive judgment that, under this kind of hypnotic suggestion, one no longer feels cold. But I suspect that judgment is harder to maintain when considering the sort of conditions that typically feature in internalist and subjectivist norms. Take for example (EN) and the Subjective Act Consequentialist Norm (SACN). In both

---

<sup>106</sup> Here, Schwitzgebel's (2011) claim that our self-knowledge is indeed worse than our knowledge of the external world is highly relevant. See section 1.

cases, the triggering condition is a probability measure on one's evidence: *that p has a probability of at least 0.7 on one's evidence* and *that  $\varphi$ -ing has the greatest expected goodness on one's evidence*, respectively. Is it possible for these conditions to be non-transparent in non-liminal cases? Suppose that S is in a central case of a proposition *q* having at least a 0.7 probability on her evidence: *q* has a 0.95 probability. Is it possible for her to be hypnotised such that, despite her best efforts, she systematically confuses central cases of high probability on her evidence for non-high probability? I'm again inclined to say yes; indeed, I'm even more inclined to say yes to this than to the analogous question for the condition of feeling cold. It would be strange indeed to say that in the imagined case, the probability of *q* on S's evidence shifts down in virtue of S's hypnosis, from a high probability to a non-high probability. Saying so would be flatly inconsistent with an objectivised understanding of the evidential probability relations. But it would also likely be inconsistent with more subjectivised understandings of evidential probability; for it seems coherent to imagine a subject being primed into wildly mistaking her own subjective priors.

Paradigmatic internalist and subjectivist norms might very well turn out not to be cosy-transparent after all, just as they might very well turn out not to possess a greater degree of contextual transparency than their rivals. Thus it remains unclear how much comfort internalists and subjectivists should take in the thought that their desire for transparency might be satiated by some related property.

## 5. Resolving the dilemma

In his postscript to "Moral Luck", Williams wrote that "what is not in the domain of the self is not in its control, and so is subject to luck" (1981a, 20). Williams' Cartesian identification of the realm of the self with the realm of agential control offered both ethicists and epistemologists a safe harbour from normative luck. But

if Anti-Cartesianism is correct, the harbour is hardly safe. Retreat into the internal or subjective will not eradicate the possibility of blameless violation. All normative theories are thus open to the objection that they drive a wedge between the normativity and culpability perspectives.

In the final two sections I offer some thoughts on how we might respond to this dilemma.<sup>107</sup> In section 5.1 I discuss what I take to be the best prospect for defending normative transparency, namely embracing a certain sort of anti-realism about the metanormative. In section 5.2 I consider what it would mean to accept the misalignment of the normativity and culpability perspectives.

### 5.1 Cleaving to Cartesianism

One philosopher's *modus ponens* is another's *modus tollens*. There *must* be transparent norms, we might think, and so there must be at least *some* transparent conditions.

The challenge for those who wish to grasp this horn of the dilemma is not merely to show that there exists some condition that constitutes a counterexample to Anti-Cartesianism. They must show that there is such a counterexample that moreover plausibly features as a triggering condition in a norm. After all, even if the condition *that one feels like  $\psi$ -ing* were transparent, this would hardly be a source of hope for the proponent of normative transparency. Few ethicists or epistemologists would want to endorse the view that one should simply act or believe as one feels like doing.<sup>108</sup> We might say more generally that the more plausible it is that a condition is transparent, the less plausible it is that it could feature as a triggering condition in a norm. A similar problem arises for the hope

---

<sup>107</sup> I don't mean these possible resolutions to be exhaustive. In particular I don't consider the possibility of simply denying the intuitive claim that non-culpable ignorance always exculpates. This would be to deny the existence of a normativity-culpability gap to begin with.

<sup>108</sup> Thanks to Krister Bykvist for emphasising this point.

to locate normative transparency in hyper-precise phenomenal conditions. Since, presumably, our natural language lacks hyper-precise phenomenal concepts, it's unclear how we could use such conditions in the formulation of norms.<sup>109</sup>

In Essay One, I discussed the possibility that phenomenal conditions that enjoy a constitutive connection to the doxastic might constitute a counterexample to Williamson's anti-luminosity thesis. In the end I argued that it's plausible that even such conditions are non-luminous (and thus non-transparent). Nonetheless, I take it that such conditions – if they exist – remain the most plausible exceptions to Anti-Cartesianism. Consider a condition C of which the following were true:

(CON): One is in C iff after careful deliberation one would come to believe one is in C; and one is not in C iff after careful deliberation one would come to believe one is not in C

Assuming that one is always in a position to carefully deliberate about whether one is in C, (CON) implies that one is in C only if one is in a position to believe that one is in C, and that one is not in C only if one is in a position to believe that one is not in C. So it seems plausible that one is always in a position to know whether one is in C.

The question is: is there any condition that is constitutively connected to the doxastic in the way that (CON) describes *and* that could plausibly feature in a norm? As discussed in Essay One, some philosophers of mind hold that simple phenomenal conditions – feeling cold, being in pain, having a desire to  $\psi$  – enjoy the kind of constitutive connection described by (CON).<sup>110</sup> But it doesn't seem that such conditions could function as triggering conditions in plausible norms. A

---

<sup>109</sup> See Hawthorne 2005 and Williamson 2005 for a discussion of the relationship between luminosity and precision in our phenomenal concepts.

<sup>110</sup> See e.g. Shoemaker 1986; Wright 1989, 1998; Heal 2001; Chalmers 2003; Coliva 2009; and various essays in Coliva 2012.

moral or epistemic norm whose triggering condition was *that one feels cold* seems absurd, while a moral or epistemic norm whose triggering condition was *that one is in pain* or *that one desires to  $\psi$*  would generate a normative theory either too masochistic or too self-indulgent.

A more promising way to use (CON) for the purposes of normative transparency is to endorse a certain form of metanormative anti-realism, namely agent-relative constructivism.<sup>111</sup> Take the following claim about the relationship between what reasons one has and the beliefs one has about one's reasons:

(REASON CON): S has reason to  $\phi$  iff after careful deliberation she would come to believe she has reason to  $\phi$ ; and S does not have reason to  $\phi$  iff after careful deliberation she would come to believe she does not have reason to  $\phi$

(REASON CON) is analogous to (CON). Thus, if (REASON CON) is true, then it seems plausible that one is always in a position to know whether one has a reason to perform any given action. If so, then it follows that the following norm is transparent:

REASON NORM (ReN): S ought to  $\phi$  iff S has reason to  $\phi$ <sup>112</sup>

If one is always in a position to know whether one has reason to perform any given action, one is always in a position to know whether one is conforming to (ReN). Thus one can never violate (ReN) through bad luck. So long as one acts in accordance with what one, after careful deliberation, would believe oneself to have reason to do, one is acting in conformity with (ReN). The possibility of

---

<sup>111</sup> See e.g. Korsgaard 1996 and Street 2008.

<sup>112</sup> I'm ignoring complications to do with competing reasons. One could substitute 'overall reason' for 'reason' in both (REASON CON) and (ReN) to address this worry.

blameless violation is eradicated. The culpability and normativity perspectives are brought into alignment.

This result shouldn't come as a surprise. Many theorists who favour constructivist views of the metanormative do so precisely because they are concerned to vindicate a certain intimacy between the normative facts and agents' beliefs about the normative facts. So conditions that are constitutively connected to the doxastic seem a natural safe harbour against the implications of Anti-Cartesianism.

Of course, the attractiveness of such a harbour will turn on one's particular metanormative inclinations. Not even all metanormative non-cognitivists will be happy to take shelter here. Those, for example, who endorse social constructivism will perhaps want to deny that individuals are always in a position to know whether they are in conformity with the socially constructed norms.<sup>113</sup> And similarly for many fans of naturalist non-cognitivism, relativism and idealised response-dependence. Only those anti-realists who embrace the Dummettian view that knowledge in a domain cannot be outstripped by truth in that domain will have the metanormative resources available plausibly to deny Anti-Cartesianism.

I won't here rehearse my reasons for thinking that even constitutively connected conditions of the type posited by (REASON CON) aren't transparent.<sup>114</sup> But I want to note that, even granting the transparency of such conditions, there are reasons to worry about this resolution of the dilemma. Earlier I said that the more plausible a given condition is as a counterexample to Anti-Cartesianism, the less plausible it is as a triggering condition in a norm. It's important to note that constructivism doesn't sidestep this dilemma. Rather, it engages it by embracing a strongly permissive view of the normative. According to the constructivist proposal, what one ought to do is a function of what, under careful deliberation,

---

<sup>113</sup> For examples of social constructivism, see Rawls 1971 and Scanlon 1998.

<sup>114</sup> See Essay One, section 5 for a discussion.

one would take one's reasons to be. This rules in the possibility of an agent who, because of his initial fund of desires, motivations and beliefs, carefully deliberates his way to moral atrocity or epistemic absurdity. The possibility here is not merely psychological, not merely that agents can end up believing that they have strong reason to commit genocide or believe patent falsities. That psychological possibility is live regardless of one's metanormative theory. Rather the possibility, on the constructivist proposal, is normative: it might very well turn out that agents *truly* believe that they have strong reason to commit genocide or believe patent falsities.<sup>115</sup> For many this will no doubt be too high a price to pay for normative transparency – and perhaps even a *reductio* against the internalist and subjectivist longing to align the culpability and normativity perspectives.

## 5.2 Accepting the normativity-culpability gap

Perhaps the commonplace desire to eradicate the phenomenon of blameless norm-violation should be treated as a symptom rather than as a starting point. A symptom of what? An undue fixation, we might say, with the centrality of the culpability perspective. In actual normative practice after all, our judgments of culpability are often of only secondary significance.<sup>116</sup> I promise to pick you up at the airport and, having lost all memory of making the promise, fail to do so.<sup>117</sup> Perhaps my ignorance of the promise excuses me. No matter. The most important thing, morally speaking, is that I failed to live up to my promise to you, and left you stranded at the airport. Or imagine that I hallucinate a talking horse and come to believe that some horses can talk. Again, perhaps my unlucky circumstances excuse my belief. But the most important thing, epistemically speaking, is that I am now encumbered with a false belief from a faulty source. In

---

<sup>115</sup> Gibbard (1999, 145) formulates this as the problem of the 'perfectly rational Caligula'. Korsgaard (1996) denies that such a case is possible on her Kantian constructivism. Street (m.s.) accepts its possibility, but denies its problematicity.

<sup>116</sup> Williams makes a similar point in his argument that we share more in common with the tragic perspective of the Greeks than we tend to think (1993b).

<sup>117</sup> This case was suggested to me by Stewart Cohen.

both cases, I have not *done well*. And doing well, unlike doing one's best, requires the cooperation of an often unforgiving world.

Breaking a promise to pick you up at the airport or hallucinating a talking horse are fairly mundane cases. But we only have to turn ourselves to more serious possibilities to see how very little the culpability perspective sometimes matters. Consider a political leader who, doing his best to act on his evidence and follow his conscience, leads his country into an illegal, immoral and materially devastating war. That he is in some sense excused for so doing is and should be of little interest to those who are not his biographer. Indeed, a first-personal fixation on one's own culpability seems to indulge a narcissism deeply at odds with the normative. It is fine and important to tell children that all that matters is that they *try their best*. But when this thought becomes a complete consolation for failure – when I, harming you, care only that I am trying my best not to – I have abandoned the normativity perspective. Williams launches a similar criticism against the Christian outlook when he says that it “associates morality simultaneously with benevolence, self-denial, and inner directedness or guilt (shame before God or oneself)” (1981b, 244). Seeking to avoid blame, moral or epistemic, is a form of cowardice, a turning inward of the normative perspective. One should not seek to be free from blame; one should seek to make things go best.

The culpability perspective has a strong grip on contemporary normative philosophy, and contemporary Western culture more generally. Perhaps this is a result both of overconfidence in our theoretical powers and naïveté about the place of luck in human life. Perhaps too it is a hangover from a certain Christian way of thinking about things, as Williams suggests: the conviction that a just God

will judge us by the quality of our souls, not the accidental results of our actions.<sup>118</sup> Certainly the Ancient Greeks, excepting perhaps Plato, were more open-eyed about the role of luck in the normative sphere.<sup>119</sup> In the *Nicomachean Ethics*, Aristotle tells us that the chief good consists in activity of the soul in accordance with *aretê* (1925, 1098a17-20). But he then tells us that no one who is not already practiced in good habits has anything to learn from this definition (1095b4-5). *Eudaimonia* “needs the external goods as well; for it is impossible, or not easy, to do noble acts without proper equipment” (1099a31-33). To be good is to do well, and to do well requires good fortune.

Our contemporary fixation on culpability thus contrasts with the tragic outlook of the Greeks – a view, as Williams says, that maintains “that what is great is fragile and what is necessary may be destructive” (1981b, 253). Indeed, so deep is our obsession with culpability that we stubbornly insist on reading the obsession into Greek tragedy. Thus the classicist E.R. Dodds (1966) laments what he calls the “moralizing interpretation” that critics and students are often keen to give *Oedipus Rex*. According to a common reading, Oedipus’ downfall is due to a moral failing on his part – his pride or overconfidence, perhaps, a ‘fatal flaw’. Not only is that reading unwarranted by Sophocles’ text, Dodds argues, but it is also unsupported by Aristotle’s use of *Oedipus* as a canonical example of tragedy. Aristotelian *hamartia* (literally, ‘missing of the mark’) – of which Oedipus, along with Thyestes, is the central example for both Aristotle (*Poetics* 13) and Plato (*Laws* 838c) – should not be interpreted as ‘fatal moral flaw’ but rather as ignorance of a non-moral fact that leads to wrongdoing. In Oedipus’ case, this is ignorance of the identity of his parents. In Thyestes’, it is ignorance that what seems to be butcher’s meat is actually the flesh of his children. The natural moral

---

<sup>118</sup> Relevant here is Ted Sider’s claim that an epistemic treatment of vagueness is incompatible with a traditional Christian picture of the afterlife (2002); Sider argues that it would be *unjust* for unknowable facts to make a difference as to whether one went to hell or heaven.

<sup>119</sup> For a discussion of Ancient Greek views on moral luck, see Nussbaum 1986. Nussbaum defends the view that Aristotelian ethics represents a resurgence of the pre-Platonic, tragic view of moral luck as an endemic fact of human existence, which even reason cannot surmount.

laws prohibit parricide, incest and cannibalism. But these laws are non-transparent. One can violate them unknowingly, through bad luck. The tragedy of Oedipus, then, is not of a great man brought down by moral weakness. It is rather of a great man brought down (to put it prosaically) by normative non-transparency.

Like us, the Ancient Greeks recognised a distinction between those moral harms performed by bad luck and those performed under agential control; Oedipus would have been acquitted by an Athenian court of murder. Nonetheless, the all-important moral state of *pollution* is one of which no court can acquit Oedipus nor Thyestes:

Thyestes and Oedipus are both of them men who violated the most sacred of Nature's laws and thus incurred the most horrible of all pollutions; but they both did so without πονηρία [*poneria*, vice], for they knew not what they did – in Aristotle's quasi-legal terminology, it was a ἁμάρτημα [*hamarthma*], not an ἀδίκημα [*adikhma*, moral injustice]. That is why they were in his view especially suitable subjects for tragedy. Had they acted knowingly, they would have been inhuman monsters, and we could not have felt for them that pity which tragedy ought to produce. As it is, we feel both pity, for the fragile estate of man, and terror, for a world whose laws we do not understand. (Dodds 1966, 39 – 40).

Perhaps it is time to give up our fixation on culpability, and return to a more tragic outlook.<sup>120</sup> The growing discomfort with the Cartesian view of the phenomenal is one reason for such a return. But another is itself normative. The impulse to eradicate luck from the normative sphere often issues, I suspect, from a laudable discomfort with the role that luck plays in other spheres of human life, and especially its intimate connection with economic and political inequality. Understandably, then, we hope to establish that *one* realm of human life is immune from luck. But as Marx said of religion, this philosophical longing might in the end be morally neutering: a revolt against injustice in theory, but not in

---

<sup>120</sup> In his (1993a) Williams might be read as making a similar suggestion in his exhortation to replace morality with ethics.

practice. In the place of such a project, we might instead have a post-Cartesian normative theory: more open-eyed, more tragic, more attentive to the fragility of our human estate.

## Acknowledgments

Special thanks to John Hawthorne, Tim Williamson, Derek Parfit, David Wiggins, David Velleman, Nishi Shah, Paul Boghossian, Crispin Wright, Jennifer Nagel, Jane Friedman, Krister Bykvist, Kian Mintz-Woo and Robert Simpson for reading and commenting on earlier drafts. Thanks also to Corine Besson, Akeel Bilgrami, Dave Chalmers, Cian Dorr, Daniela Dover, Jeremy Goodman, Anandi Hattiangadi, Harvey Lederman, Nakul Krishna, Paul Myerscough, Andreas Mogensen, Wlodek Rabinowicz, Regina Rini, Daniel Rothschild, Jim Pryor, Jeff Russell, Fred Wilmot-Smith and Steve Yablo for helpful discussion of these issues. This paper draws on a paper I co-authored with John Hawthorne, "Disagreement Without Transparency: Some Bleak Thoughts" (2013), in Christensen D. and J. Lackey, (eds.), *The Epistemology of Disagreement* (Oxford: OUP). Earlier versions of it have been presented at the All Souls Epistemology Workshop, Oxford (2013); the DPhil Dissertation Seminar, Oxford (2013); the Objective and Subjective Oughts seminar, Oxford (2013); the Epistemology of Philosophy Conference, University of Cologne (2011); the Mind and World Conference, IUC-Dubrovnik (2011); and the Oxford Philosophy Faculty women's event (2011). Many thanks to everyone in those audiences as well.

## Works Cited

- Alston, W. (1988). "An Internalist Externalism". *Synthese*, 74: 265-83. Reprinted in Alston, W. (1989), 227-45. [Page references are from reprint].
- Alston, W. (1989). *Epistemic Justification*. Ithaca: Cornell University Press.
- Alston, W. (2005). *Beyond "Justification": Dimensions of Epistemic Evaluation*. Ithaca: Cornell University Press.
- Aristotle (1925). *Nicomachean Ethics*, W.D. Ross (transl.) Oxford: Oxford University Press.
- Audi, R. (2001). "An internalist theory of normative grounds". *Philosophical Topics*, 29: 19 - 46.
- Austin, J.L. (1956). "A Plea for Excuses". *Proceedings of the Aristotelian Society*, 57: 1-30.
- Bales, R.E. (1971). "Act-utilitarianism: account of right-making characteristics or decision making procedure?" *American Philosophical Quarterly*, 8: 257-265.
- Bennett, J. (1990). "Why is Belief Involuntary?" *Analysis*, 50: 87-107.

- Bennett, J. (1998). *The Act Itself*. Oxford: Oxford University Press.
- Boghossian, P. (1989). "The Rule-Following Considerations". *Mind*, 98: 507-549.
- Boghossian, P. (1994) "Transparency of Mental Content". *Philosophical Perspectives*, 8: 33-50.
- BonJour, L. (1980). "Externalist Theories of Empirical Knowledge". *Midwest Studies in Philosophy*, 5: 53-73.
- BonJour, L. (1985). *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- BonJour, L. (2003). "A Version of Internalist Foundationalism", in BonJour, L. and Sosa, E. (eds.), *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. Malden, MA: Blackwell, 3-96.
- Broome, J. (1991). *Weighing Goods*. Oxford: Blackwell.
- Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Chalmers, D. (2003). "The Content and Epistemology of Phenomenal Belief", in Smith, Q. and Jolic, A. (eds.), *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press, 220 – 271.
- Chisholm, R.M. (1977). *Theory of Knowledge*, 2<sup>nd</sup> edn. Englewood Cliffs, NJ: Prentice-Hall.
- Coliva, A. (2009). "Self-knowledge and Commitments". *Synthese*, 171(3): 365 – 75.
- Coliva, A., (ed.) (2012). *Self and Self-Knowledge*. Oxford: Oxford University Press.
- Conee, E. and Feldman, R. (2001). "Internalism Defended". *American Philosophical Quarterly*, 38(1): 1-18.
- Conee, E. and Feldman, R. (2011). "Replies" in Dougherty, T. (ed.), *Evidentialism and its Discontents*. Oxford: Oxford University Press, 281-232.
- Cook, J.T. (1987). 'Deciding to Believe without Self-Deception' / *Journal of Philosophy*, 84: 441-46.
- Cruz, J. and Pollock, J. (1999). *Contemporary Theories of Knowledge*, 2nd edition. Lanham, MD: Rowman & Littlefield.
- Dodds, E.R. (1966). "On Misunderstanding the 'Oedipus Rex'". *Greece & Rome*, Second Series, 13(1): 37 – 49.
- Dougherty, T. (forthcoming). "Vague Value". *Philosophy and Phenomenological Research*.
- Ewing, A.C. (1947). *The Definition of Good*. New York: MacMillan.

- Farrelly, C. (2007). "Justice in Ideal Theory: A Refutation". *Political Studies*, 55(4): 844-864.
- Feldman, R. (1988). "Epistemic Obligations". *Philosophical Perspectives*, 2: 235-56.
- Foley, R. (2001). *Intellectual Trust in Oneself and Others*. Cambridge: Cambridge University Press.
- Fumerton, R. (1995). *Metaepistemology and Skepticism*. Lanham, MD: Rowman & Littlefield.
- Gibbard, A. (1990). *Wise choices, apt feelings*. Oxford: Oxford University Press.
- Gibbard, A. (1999) "Morality as Consistency in Living: Korsgaard's Kantian Lectures". *Ethics*, 110: 140-164.
- Gibbard, A. (2005). "Truth and Correct Belief". *Philosophical Issues*, 15: 338-50.
- Ginet, C. (1975). *Knowledge, Perception, and Memory*. Dordrecht: Reidel.
- Goldman, A. (1980). "The Internalist Conception of Justification". *Midwest Studies in Philosophy*, 5: 27-51.
- Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- Goldman, A. (1999). "Internalism Exposed". *Journal of Philosophy*, 96: 271-93.
- Goodman, J. (2013). "Inexact Knowledge without Improbable Knowing". *Inquiry*, 56 (1): 29-51.
- Gopnik, A. (1983). "How we can know our minds: The illusion of first person knowledge of intentionality". *Brain and Behavioral Science*, 16: 1-14.
- Griffin, J. (1996). *Value Judgement*. Oxford: Clarendon Press.
- Gruzalski, B. (1981). "Foreseeable Consequence Utilitarianism". *Australasian Journal of Philosophy*, 59: 163-76.
- Haack, S. (1997). "The Ethics of Belief Reconsidered", in Hahn, L.E. (ed.), *The Philosophy of Roderick Chisholm*. La Salle, IL: Open Court, 129-44.
- Hare, R.M. (1981). *Moral thinking: its levels, method and point*. Oxford: Clarendon Press.
- Harman, E. (2011). "Does Moral Ignorance Exculpate?" *Ratio*, 24(4): 443-468.
- Hawthorne, J. (2005). "Knowledge and Evidence". *Philosophy and Phenomenological Research*, 70(2): 452-458.
- Heal, J. (2001). "On first-person authority". *Proceedings of the Aristotelian Society*, CII: 1-19.
- Heil, J. (1984). "Doxastic Incontinence". *Mind*, 93: 56-70.

- Henriksen, R. (1995). "Unknowable Obligations". *Utilitas*, 7(2): 247-271.
- Hooker, B. (2000). *Ideal Code, Real World: a rule-consequentialist theory of morality*. Oxford: Oxford University Press.
- Howard-Snyder, F. (1997). "Rejection of Objective Consequentialism". *Utilitas*, 9: 241–248.
- Hudson, J.L. (1989). "Subjectivization in ethics". *American Philosophical Quarterly*, 26: 221–229.
- Hume, D. (1739/1978). *A Treatise of Human Nature*. L. A. Selby-Bigge (ed.), 2nd ed. revised by P.H. Niddich. Oxford: Clarendon Press.
- Hursthouse, R. (1999). *On Virtue Ethics*. Oxford: Oxford University Press.
- Jackson, F. (1991). "Decision-Theoretic Consequentialism and the Nearest and Dearest Objection". *Ethics*, 101: 461-82.
- Kagan, S. (1998). *Normative Ethics*. Boulder: Westview Press.
- Kant, I. (1785/1959). *Foundations of the Metaphysics of Morals*. L. W. Beck, (trans.). Indianapolis: Bobbs-Merrill.
- Kornblith, H. (1982). "The Psychological Turn". *Australasian Journal of Philosophy*, 60: 238-53.
- Korsgaard, C. (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.
- Kripke, S. (1982). *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
- Lasonen-Aarnio, M. (2011). "Unreasonable Knowledge". *Philosophical Perspectives*, 24(1): 1-21.
- Lenman, J. (2000). "Consequentialism and cluelessness". *Philosophy and Public Affairs*, 29: 342–70.
- Lockhart, T. (2000). *Moral uncertainty and its consequences*. Oxford: Oxford University Press.
- Mackie, J.L. (1976). *Problems from Locke*. Oxford: Oxford University Press.
- Mason, E. (2003). "Consequentialism and the 'Ought Implies Can't' Principle". *American Philosophical Quarterly*, 40(4): 319–331.
- Mill, J.S. (1861/2001). *Utilitarianism*, 2nd edition. George Sher (ed.) Indianapolis: Hackett.
- Montmarquet, J. (1993). *Epistemic Virtue and Doxastic Responsibility*. Lanham, MD: Rowman & Littlefield.
- Moore, G.E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.

- Moser, P.K. (1989). *Knowledge and Evidence*. New York: Cambridge University Press.
- Nagel, T. (1979). *Mortal Questions*. New York: Cambridge University Press.
- Nagel, T. (1989). *The View from Nowhere*. Oxford: Oxford University Press.
- Naylor, M.B. (1985). "Voluntary Belief". *Philosophy and Phenomenological Research*, 45: 427-36.
- Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Belknap Press.
- Nozick, R. (1993). *The Nature of Rationality*. Princeton: Princeton University Press.
- Nussbaum, M. (1986). *The Fragility of Goodness*. Cambridge: Cambridge University Press.
- Oddie, G. (1995). "Moral uncertainty and human embryo experimentation," in Fulford, K.W.M., Gillett, G. and Soskice, J.M. (eds.), *Medicine and Moral Reasoning*. Oxford: Oxford University Press.
- Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.
- Parfit, D. (2011). *On What Matters*. Oxford: Oxford University Press.
- Pettit, P. and Brennan, G. (1986). "Restrictive consequentialism". *Australasian Journal of Philosophy*, 64:438-55.
- Plantinga, A. (1988). "Positive Epistemic Status and Proper Function". *Philosophical Perspectives*, 2: 1-50.
- Plantinga, A. (1993a). *Warrant: The Current Debate*. New York: Oxford University Press.
- Plantinga, A. (1993b). *Warrant and Proper Function*. New York: Oxford University Press.
- Pollock, J. and Cruz, J. (1999). *Contemporary Theories of Knowledge*, 2<sup>nd</sup> edn. Totowa, NJ: Rowman & Littlefield.
- Prichard, H.A. (1932/1949). "Duty and Ignorance of Fact," in Ross, W.D. (ed.), *Moral Obligation*. Oxford: Clarendon Press, 18-39.
- Pritchard, D. (2004). "Epistemic Luck". *Journal of Philosophical Research*, 29: 193-222.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford: Oxford University Press.
- Pritchard, D. (2012). "Anti-Luck Virtue Epistemology". *Journal of Philosophy*, 109: 247-79.
- Pryor, J. (2001). "Highlights of Recent Epistemology". *British Journal of Philosophy of Science*, 52: 95 - 124.
- Radcliffe, D. (1997). "Scott-Kakures on Believing at Will". *Philosophy and Phenomenological Research*, 57: 145-51.

- Railton, P. (1984). "Alienation, consequentialism, and the demands of morality". *Philosophy and Public Affairs*, 13: 134–71.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rosen, G. (2004). "Skepticism About Moral Responsibility". *Philosophical Perspectives*, 18: 295-313
- Ross, W.D. (1939). *Foundations of Ethics*. Oxford: Oxford University Press.
- Scanlon, T.M. (1998). *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Scanlon, T.M. (2001). "Thomson on Self-Defense", in Byrne, A., Stalnaker, R. and Wedgwood, R. (eds.), *Fact and Value: Essays on Ethics and Metaphysics for Judith Jarvis Thomson*. Cambridge, MA: MIT Press, 199–215.
- Scanlon, T.M. (2008). *Moral Dimensions: Meaning, Permissibility, and Blame*. Cambridge, MA: Harvard University Press.
- Schwitzgebel, E. (2008). "The Unreliability of Naïve Introspection". *Philosophical Review*, 117: 245-273.
- Schwitzgebel, E. (2011). *Perplexities of Consciousness*. Cambridge, MA: MIT Press.
- Scott-Kakures, D. (1994). "On Belief and Captivity of the Will". *Philosophy and Phenomenological Research*, 54: 317-33.
- Sepielli, A. (2009). "What to do when you don't know what to do", in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics IV*. Oxford: Clarendon Press, 5-28.
- Shapiro, S. (1998). "The Difference That Rules Make", in Bix, B. (ed.), *Analyzing Law: New Essays in Legal Theory*. Oxford: Clarendon Press.
- Shoemaker, S. (1986). *The First Person Perspective and Other Essays*. Oxford: Oxford University Press.
- Sider, T. (2002). "Hell and Vagueness". *Faith and Philosophy*, 19: 58-68.
- Sidgwick, H. (1907). *The Methods of Ethics*, 7th edition. London: Macmillan.
- Smart, J.J.C. (1961/1973). "Outline of a system of utilitarian ethics", in Smart, J.J.C. and Williams, B.A.O. (eds.), *Utilitarianism: for and against*. Cambridge: Cambridge University Press, 3–74.
- Smith, H. (1983). "Culpable Ignorance". *Philosophical Review*, 92: 543 - 571.
- Smith, H. (1988). "Making moral decisions". *Noûs*, 22: 89–108.
- Sosa, E. (1991). *Knowledge in Perspective*. Cambridge: Cambridge University Press.
- Steglich-Petersen, A. (2010). "Luck as an epistemic notion". *Synthese*, 176: 361-377.
- Steup, M. (1988). "The Deontic Conception of Epistemic Justification". *Philosophical Studies*, 53: 65-84.

- Steup, M. (1999). "A Defense of Internalism", in Pojman, L. (ed.), *Theory of Knowledge: Classical and Contemporary Readings*. Belmont, CA: Wadsworth, 373-84.
- Street, S. (2008). "Constructivism about Reasons", in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics*, vol. 3. Oxford: Clarendon Press, 207-245.
- Street, S. (m.s.). "Objectivity and Truth: You'd Better Rethink It". Available at: [https://files.nyu.edu/ss194/public/sharonstreet/Writing\\_files/Paper%2012%20for%20website%20-%20Objectivity%20and%20Truth%20-%20You'd%20Better%20Rethink%20It.pdf](https://files.nyu.edu/ss194/public/sharonstreet/Writing_files/Paper%2012%20for%20website%20-%20Objectivity%20and%20Truth%20-%20You'd%20Better%20Rethink%20It.pdf)
- Suikkanen, J. (2008). "Williamson and (Moral) Realism". PEA Soup blog. Posted 27th March, 2008. URL: <http://peasoup.typepad.com/peasoup/2008/03/williamson-and.html#more>
- Thomson, J. J. (2008). *Normativity*. Chicago and La Salle, IL: Open Court Press.
- Timmons, M. (2012). *Conduct and Character*, 6th ed. Wadsworth: Cengage Learning.
- Toulmin, S.E. (2001). *Return to Reason*. Cambridge, MA: Harvard University Press.
- Wedgwood, R. (2002). "Internalism Explained". *Philosophy and Phenomenological Research*, 65(2): 349-369.
- Wedgwood, R. (2003). "Choosing Rationally and Choosing Correctly", in Stroud, S. and Topple, C. (eds.), *Weakness of Will and Practical Irrationality*. Oxford: Oxford University Press, 201-230.
- Wedgwood, R. (2010). "There are unknowable moral truths". PEA Soup blog. Posted 10th August 2010. URL: <http://peasoup.typepad.com/peasoup/2010/08/there-are-unknowable-moral-truths.html>
- Williams, B. (1976). "Moral Luck". *Proceedings of the Aristotelian Society*, 50: 115-135.
- Williams, B. (1981a). *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, B. (1981b). "Philosophy", in Finley, M.I. (ed.), *The Legacy of Greece: A New Appraisal*. Oxford: Clarendon Press, 202-55.
- Williams, B. (1993a). "Postscript", in Staman, D. (ed.), *Moral Luck*. Albany: State University of New York Press, 251-8.
- Williams, B. (1993b). *Shame and Necessity*. Berkeley: University of California Press.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, T. (2005). "Replies to Commentators". *Philosophy and Phenomenological Research*, 70(2): 468-491.
- Williamson, T. (2007). "On Being Justified in One's Head", in Timmons, M., Greco, J. and Mele, A. (eds.), *Rationality and the Good: Critical Essays on the Ethics and Epistemology of Robert Audi*. Oxford: Oxford University Press.

Williamson, T. (2011). "Improbable Knowing", in Dougherty, T. (ed.), *Evidentialism and its Discontents*. Oxford: Oxford University Press, ch. 9.

Wittgenstein, L. (1953). *Philosophical Investigations*. G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.). Oxford: Blackwell.

Wright, C. (1989). "Wittgenstein's later philosophy of mind: Sensation, privacy and intentions". *Journal of Philosophy*, 86(11): 622-34.

Wright, C. (1998). "Self-knowledge: the Wittgensteinian legacy", in Wright, C., Smith, B. and Macdonald, C. (eds.), *Knowing Our Own Minds*. Oxford: Clarendon Press, 13-45.

Zagzebski, L. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.

Zimmerman, M. (1996). *The Concept of Moral Obligation*. Cambridge: Cambridge University Press.

Zimmerman, M. (1997). "Moral Responsibility and Ignorance". *Ethics*, 107: 410 – 428.

Essay Three

The Archimedean Urge

## I. Genealogical Contingency

Scepticism about philosophy is as old as philosophy itself. In Aristophanes' *The Clouds*, Socrates orders his hapless student Strepsiades to lie down on a couch in order to make him more receptive to philosophical inspiration. Instead Socrates catches Strepsiades masturbating under the bedclothes. Aristophanes' suggestion is that it amounts to much the same thing.

Like philosophy itself, scepticism about philosophy has its modes and fashions. Sometimes the accusation, as with Aristophanes, is that a seemingly lofty activity is in fact chicanery, nonsense in the service of all too human desire. According to diagnostic scepticism, as with the later Wittgenstein, philosophy is charged with being a symptom of a pathology or confusion. Scientistic scepticism impugns philosophy for falling short of some putative standard met by respectable – that is, empirical – modes of enquiry. A fourth variety of scepticism complains that the building blocks of philosophy – the judgments and concepts on which it all hangs – are contingent features of whoever it is doing the philosophising: his or her history, culture, character, language, education, neurophysiology.

This last kind of worry, at least in its most general form as a worry about the *genealogical contingency* of thought, has gripped much of the Western humanistic tradition since the 19th century, the moment of the so-called historicist turn. Though he was not the first to suggest it – we see similar rustlings in Vico and Herder in the preceding century – it was Hegel who really taught Western intellectual culture to think historically about thought itself: to see thought as inseparable from its temporal conditions. As “each child is in any case a *child of his time*...thus, philosophy, too, is *its own time comprehended* in thoughts” (1820/1991, 21). And yet, Hegel's historicism was not of the worrying kind.

Though thought unfolds historically, it is hardly, for Hegel, the mere product of contingency:

[Spirit] has the History of the World for its theatre, its possession, and the sphere of its realization. It is not of such a nature as to be tossed to and fro amid the superficial play of accidents, but is rather the absolute arbiter of things; entirely unmoved by contingencies, which, indeed, it applies and manages for its own purposes (1837/2001, 70).

Thus Hegel's genealogy of thought is, in Bernard Williams' terms, *vindictory* rather than *shameful*, confirming rather than undermining the authority of human reason. It was Nietzsche who taught Western culture to think *anxiously* about thought's genealogy – to worry that genealogical excavation might reveal its *pudenda origo* (Nietzsche 1881/1911).<sup>121,122</sup>

From poststructuralists like Lacan and Foucault to naturalistic debunkers of religion like Richard Dawkins and Christopher Hitchens, Nietzsche has many heirs in dwelling on the supposedly shameful implications of genealogical contingency. Indeed much of the intellectual history of the 20<sup>th</sup> century can be told in terms of a war between those committed to genealogical excavation and those, like Karl Popper and Leo Strauss, who thought such excavation both alethically irrelevant and morally pernicious. For the most part analytic philosophy has been on the side that thinks there not much point to revelations of genealogical contingency. After all, we believe everything we do because of

---

<sup>121</sup> Again, Nietzsche wasn't the first to think this way. Williams argues that in Book 2 of *Republic*, Plato has Socrates reject the genealogical accounts of justice put forward by Glaucon and Adeimantus because he thinks they would undermine the intrinsic value of justice (Williams 2002). And we of course see similar sceptical tendencies in Hume's writings on religion. Xenophanes of Kolophon is perhaps the most amusing early example of genealogical unmasking:

Mortals suppose that the gods are born (as they themselves are), and that they wear man's clothing and have human voice and body. But if cattle or lions had hands, so as to paint with their hands and produce works of art as men do, they would paint their gods and give them bodies in form like their own—horses like horses, cattle like cattle (1898, fragments 5-6).

<sup>122</sup> Though Nietzsche, ever cryptic, also wrote: "The origin becomes of less significance in proportion as we acquire insight into it; whilst things nearest to ourselves, around and within us, gradually begin to manifest their wealth of colours, beauties, enigmas, and diversity of meaning, of which earlier humanity never dreamed" (1881/1911 I:44).

contingent facts about ourselves: that we exist, that we were born where and when we were, that we received the education we did. The mere fact that a belief or practice has a contingent origin does not entail that it is false or unjustified; to think otherwise is to commit the genetic fallacy.

Worries about genealogical contingency are nonetheless not entirely absent from analytic philosophy. Indeed they are increasingly common. Many ethicists claim that the evolutionary origins of our moral judgments demand that we abandon or revise those judgments – or at least adopt an anti-realist construal of their contents (Harman 1977, 1986; Singer 1981, 2005; Ruse 1985; Ruse and Wilson 1986; Gibbard 1990; Kitcher 2005, 2011; Joyce 2006, ch. 6; Street 2006, 2008a, 2011; Greene 2008; Huemer 2008; Rosenberg 2011.)<sup>123</sup> James Ladyman and Don Ross (2007) argue that the evolutionary origins of our metaphysical judgments should make us suspicious of their reliability. Genealogically-motivated arguments have also been made in support of anti-realism in math (Benacerraf 1973; Field 1989) and logic (Cooper 2003), and the falsity of both theism (Dennett 2006)<sup>124</sup> and naturalism (Plantinga 1993, ch. 12, 1994, 2002). Finally, the new sub-discipline of ‘experimental philosophy’ is largely devoted to arguing that people’s judgments about epistemology, ethics, philosophy of language and metaphysics systematically vary with culture, gender, socioeconomic status and extent of philosophical training – and thus that these judgments should be cleansed from philosophical practice (for an overview, see Knobe and Nichols 2008, forthcoming; and Alexander 2012).<sup>125,126</sup>

---

<sup>123</sup> With the same logic but in a different spirit, Thomas Nagel (2012) infers from the putative incompatibility of the evolutionary origins of moral judgment with moral realism that the evolutionary explanation of our moral judgments is false.

<sup>124</sup> Genealogical attacks on theism have a particularly storied pedigree. See Feuerbach 1841; Marx 1844; Nietzsche 1887; Freud 1927.

<sup>125</sup> Not all of experimental philosophy participates in this sceptical project; much of it is focused on using empirical methods to address first-order philosophical issues, usually in philosophy of mind and action and moral psychology. Jesse Prinz (2008) helpfully suggests using the term ‘empirical philosophy’ for this non-sceptical part of experimental philosophy, and reserving the label ‘experimental philosophy’ for the empirically-driven critique of philosophical judgment. I will follow Prinz in his use of the terms ‘experimental philosophy’ and ‘experimental philosophers’.

What should we make of such genealogical scepticism about philosophical judgment? No doubt many find it intuitively compelling. Thus Joshua Knobe and Shaun Nichols predict that encountering the results of experimental philosophy will cause in philosophers a

...crisis akin to that of the [Christian] child confronted with religious diversity...For the discovery of religious diversity can prompt the thought that it's in some sense accidental that one happens to be raised in a Christian household rather than a Hindu household. This kind of arbitrariness can make the child wonder whether there's any reason to think that his religious beliefs are more likely to be right than those of the Hindu child...And just as some Christian children come to think that there's no rational basis for preferring Christian to Hindu beliefs, we too might come to think that there's no rational basis for preferring Western philosophical notions to Eastern ones (2008, 11).

As a descriptive matter, Knobe and Nichols might be right. But the tacit normative claim here – that philosophers *ought* to experience such a crisis – is far more vexed. Not every crisis is rational. And not every revelation of genealogical contingency undermines judgment. My judgment that Paris is the capital of France is contingent on the fact that I exist at all, that I possess the concepts *Paris* and *France*, and that I have been taught that the capital of France is Paris. And yet none of these revelations of genealogical contingency seem to undermine my claim to know that Paris is the capital of France. If they do, we have entered a realm of wholesale scepticism, in which none of my judgments are secure. Such wholesale scepticism is presumably not to the taste of most genealogical sceptics of philosophy. After all, their scepticism is based on facts that they take themselves to *know* about our philosophical judgments – namely, that they have the particular genealogies (in evolutionary history, or culture, or education) they do (cf. Bealer

---

<sup>126</sup> A related topic is the much discussed issue of *peer disagreement*. Proponents of so-called 'conciliatory' or 'split-the-difference' views on peer disagreement sometimes suggest that peer disagreement constitutes some sort of genealogical evidence: specifically, evidence about the unreliability of one's belief-forming mechanism. But not all discussions of peer disagreement treat it as an issue about genealogy. So for the most part I set it aside. However I return to it in section 3.5, where I explain how my 'Argument from Unreliability' might be recruited to support the conciliatory position. See my paper with John Hawthorne (2013) for a discussion of the disagreement debate.

1992; Foley 1998; Sosa 1998, 2005; Tidman 1996; Yablo 1993). Such sceptics must do more than reveal that our philosophical judgments are genealogically contingent. They must show why the kind of genealogical contingency exhibited by (some or all) philosophical judgments should cause a *rational* crisis of faith.

My central aim in this paper is to show that any plausible argument along these lines tacitly relies on contested first-order epistemological premises.<sup>127</sup> I will do this by constructing five arguments on behalf of the genealogical sceptic: the argument from sensitivity (3.1); from explanatory inertness (3.2); from coincidence (3.3); from probability on the evidence (3.4) and from unreliability (3.5). In each case, I will show how the soundness of the argument turns on various first-order questions in epistemology – controversial questions about explanation, evidence, justification, defeat, methods, bootstrapping and so on. This is despite the claim, commonly made by genealogical sceptics, to take an Archimedean view of philosophy – inhabiting the philosophically neutral perspectives of sociology, history, science or common sense. No such Archimedean stance is available; to argue against philosophy on the grounds of genealogical contingency requires taking a stance within epistemology, and thus within philosophy itself.<sup>128</sup>

This lack of an Archimedean stance has, I argue, two implications. First, genealogical scepticism risks self-defeat, impugning commitment to its own

---

<sup>127</sup> Knobe and Nichols' analogy between the genealogical contingency of religious belief and the genealogical contingency of philosophical judgments is here very apt, but not for the reasons they suppose. Over two decades of philosophical debate about the rationality of religious exclusivism – whether one *ought* to have a 'crisis of faith' in light of the genealogical contingency of one's religious belief – reveal just how subtle the epistemic issues in play are. See e.g. Smith 1976; Hick 1980, 1982, 1989, 2001; Gutting 1982; Plantinga 1986, 1993, 1995, 2000; Alston 1988a; Runzo 1988; Silver 2001; Willard 2001.

<sup>128</sup> The philosophical use of the metaphor of the Archimedean lever – about which, Archimedes (according to Pappus of Alexandria) famously said, "Give me a place to stand, and I will move the Earth" – has a long history. See for example Descartes 1641/1993, 17; Arendt 1958, VI. 36; Rawls 1971, 261; Berlin 1979, 114–15; Williams 1985, ch. 2; Dworkin 1996. Thanks largely to Dworkin, the notion of 'Archimedeanism' in philosophy is now usually taken to indicate a particular view in moral philosophy, namely that a meaningful distinction can be drawn between first-order normative claims and meta-ethical claims. I mean to use the notion in a broader way.

premises. For our epistemological judgments seem to be as systematically shaped by our contingent circumstances as any other of our philosophical judgments. Second, philosophers have available to them the resources required to defend against genealogical scepticism. Specifically, an epistemology that embraces the role that luck plays in the acquisition of knowledge can contribute to a metaphilosophy that is robust against genealogical attacks. For inasmuch as one takes seriously the idea that forces outside one's control can determine whether one knows, one can be sanguine about the phenomenon of *genealogical luck* – the luck of having knowledge-conducive or knowledge-hindering genealogies.

I proceed as follows. In section 2 I make some preliminary remarks about the notion of 'philosophical judgment' that I am using in the paper. In section 3 I lay out five arguments on behalf of the genealogical sceptic and show, in each case, the controversial epistemological premises on which they rely – thereby highlighting the epistemic resources available to those philosophers who wish to dismiss genealogical scepticism, and also showing why genealogical scepticism risks self-defeat. In section 4 I discuss whether the genealogical sceptic's argument can be charitably rehabilitated as an *ad hominem* or pragmatic attack. And in section 5, I conclude with some thoughts on what an Anti-Archimedean metaphilosophy might look like. Such a metaphilosophy retains a robustly realist notion of philosophy's subject matter and method, while acknowledging that getting philosophy right is partly a matter of luck.

## 2. 'Philosophical Judgments'

### 2.1 The content of philosophical judgments

Some genealogical sceptics – especially those who favour evolutionary debunking arguments against metaphysics, mathematics or moral realism – explicitly

presuppose a realist construal of the relevant subject matter. That is, they suppose that philosophical judgments are about the way the world is, independently of our beliefs about the way the world is. Whether those judgments are true or not – and whether they constitute knowledge or not – depends on whether they get the world as it really is correct. Thus Ladyman and Ross (2007) argue that the evolutionary origins of our metaphysical judgments debunk their reliability as guides to the way the world really is. And Sharon Street (2006) argues that the evolutionary origins of our moral beliefs, construed as beliefs about the mind-independent moral reality, undermine our claims to moral knowledge.

But other genealogical sceptics – especially experimental philosophers – sometimes assume that the subject matter of philosophical judgment isn't the way the world is independently of how we conceive of it, but rather merely our concepts of the world. According to this *mere conceptual analysis* model of philosophy,<sup>129</sup> philosophical theorising aims only to articulate conditions for the proper application of our concepts – e.g. *knowledge*, *freedom*, *identity* – where those conditions are taken to be constitutively tied to patterns of use by competent users of the concepts. No doubt some philosophers conceive of their enterprise in just this way.<sup>130</sup> But many do not, taking themselves instead to be theorising about, and acquiring knowledge of, the way the world really is, albeit sometimes through the use of conceptual analysis (see n. 129). That experimental philosophers often

---

<sup>129</sup> I say 'mere' conceptual analysis to distinguish this from the view that philosophy aims at knowledge of the conceptually-independent world *via* conceptual analysis. One can of course think that we can uncover truths about the world – e.g. the nature of knowledge or whether we have free will – by analysing our concepts, e.g. of knowledge or free will (cf. Williamson 2007, 14; McGinn 2012). On such a view, philosophy sometimes engages in conceptual analysis but philosophy isn't properly said to be *mere* conceptual analysis. (A similar point could be made about the conception of philosophy as *mere* semantic analysis – as opposed to the view that philosophers can come to know about the world through semantic descent.)

<sup>130</sup> Alvin Goldman for example writes that the “evidential status of...intuitions is of the constitutively grounded variety. It's part of the nature of concepts (in the personal psychological sense) that possessing a concept tends to give rise to beliefs and intuitions that accord with the contents of the concept” (2007, p. 4; see also Goldman 2010). Rawls can sometimes be read as saying something similar about his reflective equilibrium method, as when he writes: “the procedure of reflective equilibrium does not assume that there is one correct moral conception. It is, if you wish, a kind of psychology...” (1974, 7). Thanks to Tanya Goodchild for drawing my attention to both of these passages.

presuppose the ‘mere conceptual analysis’ model of philosophy, I suspect, has something to do with that model’s susceptibility to a fairly straightforward sceptical attack. If it turns out that people’s application of a concept like *freedom* varies widely, then any theory that claims to be an analysis of that concept but that only captures one pattern of use is suspect. Or, more broadly, if the goal of philosophy is to model the actual use of certain concepts, then anything less than a richly empirical approach would seem mistaken.

Of course, even granted such a mere conceptual analysis model, variation in philosophical judgments does not entail the falsity of a theory that only captures a subset of (apparent) use. For one might infer from the empirical data produced by experimental philosophers – data, they claim, showing that competent users apply the same concept differently – that *multiple* concepts are in play.<sup>131</sup> Different genealogies, one might say, yield different concepts. Alternatively, one might infer that certain genealogies result in incompetent use of concepts; judgments issuing from such incompetence should not be included in the theory.<sup>132</sup> So even on a mere conceptual analysis model, building theories on genealogically contingent, (apparently) non-universal judgments is not necessarily misguided. Nonetheless, with such a metaphilosophical conception in the background, a sceptical argument is not terribly difficult to construct.

A more interesting question – the question that is my focus here – is whether considerations of genealogical contingency undermine philosophical judgments *realistically* construed. After all, it is one thing to argue that variation in how people apply the concept *freedom* undermines monolithic theories of the concept, and quite another thing to argue that the genealogical contingency of my judgments about freedom undermines my right to use those judgments in

---

<sup>131</sup> See Sosa 2008. Of course such a response is less convincing in the case of intrapersonal rather than interpersonal variation – though the kind of semantic blindness posited by certain kinds of contextualist theories makes it perhaps more plausible.

<sup>132</sup> See Kauppinen 2007.

theorising about freedom. Whether a successful case can be made for the latter kind of claim – that genealogical contingency undermines philosophical judgments realistically construed – will be my focus here.

Both the realist and mere conceptual analysis model of philosophy's subject matter should be distinguished from an *anti-realist* conception. According to the sort of anti-realism I have in mind, truths in a particular philosophical domain are constitutively connected to our judgments about that domain. This is not because, as with the mere conceptual analysis model, the relevant philosophical subject matter is simply our concepts. Rather, it is because the objects of inquiry – say, the moral or mathematical truths – are themselves a function of our judgments about them. Some philosophers claim that construing certain domains of philosophical judgment anti-realistically insulates them from genealogical scepticism. For example, Street (2006) argues that we should favour her brand of metanormative constructivism about evaluative truths over a realist construal on precisely these grounds.

As I said, my focus here is on whether philosophical judgments realistically construed are under threat from revelations of their genealogical contingency. Thus the question of whether philosophical judgments *anti-realistically* construed are similarly under threat is largely orthogonal to my line of inquiry – though inasmuch as worries about genealogical contingency motivate a flight from realism to anti-realism, my discussion has implications for anti-realists. That said, there is one way in which the question about anti-realism is directly pertinent to my own argument. I argue that, because the genealogical sceptic must rely on epistemological premises if his argument against philosophical judgment is to be at all plausible, he risks self-defeat, undermining his commitment to his own premises. But the genealogical sceptic might retort that his epistemic judgments, *anti-realistically construed*, are not susceptible to his own genealogical scepticism

– and thus his argument is not self-defeating. Whether this might be a successful strategy turns on the particular species of anti-realism invoked and vexed questions about the normative force of anti-realist epistemic truth. As such I won't offer a general defence of my supposition that anti-realism does not constitute an escape from self-defeat.<sup>133</sup> Those genealogical sceptics who are keen to dodge the accusation of self-defeat might do well to seek solace here.

## 2.2 The scope of philosophical judgments

What is a 'philosophical judgment'? As I'm using the term – that is, freely and untheoretically – a philosophical judgment is an occurrent belief about a bit of philosophical subject matter that arises from reflection on that bit of philosophical subject matter. By 'philosophical subject matter' I just mean those things with which philosophers, *qua* philosophers, typically concern themselves: questions in metaphysics, ethics, epistemology, aesthetics and so on.<sup>134</sup>

This isn't much of an account, and both defenders and debunkers might protest against it. Experimentally-motivated sceptics typically announce their target to be the philosophical use of 'intuitions'.<sup>135</sup> Not all philosophical judgments, they might argue, are (or are the result of) such intuitions, and so not all philosophical judgments are within the scope of their attack. Specifically, those philosophical

---

<sup>133</sup> Though I discuss specific instances of that supposition. See section 4, especially n. 184.

<sup>134</sup> Thus on my view, of two judgments with the same propositional content, one might be philosophical while the other might not.

<sup>135</sup> Just what an 'intuition' is is a matter of both contention and confusion. The confusion stems from the use of 'intuition' variously to pick out a particular kind of belief (a philosophical belief rather than, say, an empirical belief), a phenomenal episode that gives rise to such a belief, and a particular kind of cognitive faculty. I find it helpful to distinguish these three by reserving the term 'intuition' for the phenomenal episode-type that gives rise to a certain class of ('intuitive') judgments, and which might be produced by a 'faculty of intuition'. What an intuition is in this sense – i.e. a kind of phenomenal episode – is a matter of contention. Some philosophers are minimalists about intuition, assimilating it to more familiar phenomenal episodes – for example, guesses or hunches (Gopnik and Schwitzgebel 1998, 78), conceivings (Chalmers 2002, 155-6), or conscious dispositions or inclinations to believe (Sosa 1996, 2007; Williamson 2007, 215ff). Anti-minimalists maintain that intuition is some sort of *sui generis* mental state, often a kind of intellectual seeing (Bealer 1992, 1998 §1; Conee 1998; Pust 2000, ch. 2; Huemer 2007 §1; Bengson m.s.) Part of my reason for avoiding talk of intuitions is a desire to sidestep this debate.

judgments on which experimental debunkers themselves rely – if I’m right, various epistemological judgments – might then fall outside their sceptical purview. While intuitions (or better, intuitive judgments – see n. 135) are spontaneous reactions to thought experiments, the sceptics might argue, the epistemic judgments they require are, instead, the result of careful deliberation, inference and, we might suppose, centuries of development of the scientific method. The upshot is that arguments for genealogical scepticism about the specific class of *intuitive* judgments is hardly wide enough to risk self-defeat.

Meanwhile, some defenders of philosophical judgment might protest, in the other direction, that the distinction between so-called ‘philosophical’ and ‘non-philosophical’ judgments is too shallow and uninteresting to serve any theoretical purpose. Williamson (2004a, 2004b, 2007, forthcoming) for example argues that there is no meaningful distinction to be drawn between those judgments that are philosophical and those that are not: our ability to make philosophical judgments is continuous with our ability to make reliable ‘real world’ judgments. Thus our judgments about hypothetical situations – as in the kind philosophers have about thought experiments – are the product of a mundane capacity for counterfactual and modal reasoning (Williamson 2004b). And our ‘philosophical’ judgments are no more spontaneous or non-inferential than many of our non-philosophical judgments, for example the scientist’s judgment that one theory fits the evidence better than the other, or my judgment that there aren’t any hobgoblins (Williamson forthcoming). The strong continuity between ‘philosophical’ and ‘non-philosophical’ judgments means, for Williamson, that any sceptical project that targets the former will invariably sink the latter. Scepticism about all judgment waits in the wings.

In such debates real neutrality is not possible. How one carves up the space matters, settling various questions in advance. But even if neutrality isn’t possible,

I hope some sort of dialectical balance is. Thus I'm going to assume that there is a genuine category of judgments that we can call 'philosophical' (*pace* Williamson), but that there is no deep distinction within that category between those judgments that are 'intuitive' and those that are not (*pace* experimental sceptics). The first assumption gives ground to the genealogical sceptic, and as such is dialectically charitable. The second assumption does not give the sceptic as much ground as he would like. So I will briefly say something in its defence.

What might make a philosophical judgment 'intuitive' rather than non-intuitive? The common answer is that intuitive judgments are arrived at spontaneously, without inference or intellectual deliberation, and that they arise in response to thought experiments. One just considers a question about a hypothetical case – *does Gettier's Jones know?* – and a compelling answer pops into one's head. Other philosophical judgments – say, whether epistemic internalism or externalism is correct, or whether classical logic should be revised to handle vagueness – are arrived at through deliberation and argument, and not in response to thought experiments. Thus intuitive judgments are available to philosophers and folk alike, while the latter kind of judgment is the special remit of philosophical deliberation.

But this distinction is naïve. For all our philosophical judgments are inferential in the sense that they rely on certain background theories, whether explicit or tacit. Our judgment that Gettier's Jones doesn't know tacitly relies on a theory that holds knowledge incompatible with certain forms of luck. Thus our judgments about cases are susceptible to penetration by theory: converts to consequentialism or externalism often start having consequentialist and externalist responses to thought experiments. Conversely, all our philosophical judgments are spontaneous in the sense that they must, on pain of regress, bottom out at a leap of doxastic compulsion. The epistemic externalist might have deliberated her way to

externalism in the sense that she thought about the choice actively and carefully, weighing up the merits and costs. But ultimately her plumping for externalism over internalism involved a compelling feeling that it was right. So the spontaneous/inferential distinction will not cut very deep.

Neither, I think, will the distinction between those philosophical judgments that arise from consideration of thought experiments and those that do not. As one reasons one's way to externalism or internalism, one will in all likelihood consider some cases along the way – about brains-in-vats, clairvoyants, and so on. Does this make one's ultimate judgment that, say, epistemic externalism is correct an intuitive judgment, given that it was based in part on judgments about cases? If so, then the class of intuitions will swamp nearly all philosophical judgment. If not, then the sceptic seems to be in the inconsistent position of telling philosophers that their intuitive judgments aren't reliable but that they can rely on them to form justified theories. Thus I will assume that, while we can talk meaningfully about the class of 'philosophical' judgments, there is no deep distinction to be made between intuitive and non-intuitive judgment within that class.

### 2.3 'Our' philosophical judgments

It is my goal in this paper to bring a diverse set of sceptical claims under a single umbrella. These claims are united by a concern with the genealogical contingency of our philosophical judgments. But they differ substantially in detail. One important way in which they differ is who is meant by the 'our' in the phrase 'our philosophical judgments'. For evolutionary debunking arguments, 'our' refers to the human species. For the arguments usually advanced by experimental philosophers, 'our' refers to analytic philosophers (as opposed to the philosophically untrained), or to Westerners (rather than Easterners), men (rather than women), people of high socioeconomic status (rather than of low

socioeconomic status), and so on. I will use ‘our’ variously to refer to all of these groups, with the relevant contrasts; which group I am referring to should be clear, I hope, from context.

A second difficulty is that, since the sceptical claims I am bringing together are diverse despite their united focus on genealogical contingency, there is no assurance from the outset that they can be usefully treated as a group. Rather than offering such an assurance, I hope to demonstrate through argument that they can. That said, sometimes my formulations of genealogical scepticism will be more amenable to some sceptical claims than others – evolutionary scepticism rather than empirical scepticism, or vice versa. In the end, though, I hope to show what I take these different claims to have deeply in common, while doing some justice to their diversity.

### 3. Arguments for Genealogical Scepticism

With those preliminaries out of the way, I will now turn to the task of laying out possible arguments for genealogical scepticism. I will offer a total of five such arguments. I don’t mean the discussion to be exhaustive, but I do outline what I take to be the most common or plausible strategies for advancing the cause of genealogical scepticism about philosophy. In each instance I hope to show how the soundness of the argument turns on subtle and contested issues in first-order epistemology.

#### 3.1 Insensitivity

To simplify matters, let us suppose that you belong to a particular group – let’s call it ‘Westerners’ – the members of which share all the same philosophical judgments. Meanwhile, the members of a distinct group – ‘Easterners’ – share all

the opposing philosophical judgments. For example, you and all your fellow Westerners share the judgment that Gettier's Jones doesn't know, while Easterners all share the judgment that Jones does know.<sup>136</sup> And you and your fellow Westerners share Kripkean judgments about reference while Easterners share non-Kripkean judgments.<sup>137</sup> And so on. It follows that for each philosophical judgment, one group is right and the other wrong.<sup>138</sup> Now suppose a sceptic argues as follows: "Take any one of your philosophical judgments, for example your judgment that Gettier's Jones doesn't know. The empirical evidence strongly suggests that people's judgments about Gettier cases are caused by their cultural upbringing rather than by the truth about those cases. This means that even if Jones *did* know (holding the other facts of the Gettier case fixed), you still would believe he didn't know!" The sceptic is in effect arguing that your Gettier judgment is *insensitive* to the truth, where S's belief that *p* is sensitive just in case, if *p* were false, S wouldn't believe *p*.<sup>139,140</sup> Thus we have:

### Argument from Insensitivity (AI)

For any philosophical judgment that *p* shared by you and your fellow Westerners:

- (1) Your judgment that *p* is insensitive to the truth of *p*
- (2) Sensitivity is a condition on knowledge
- (3) Therefore, you don't know *p*

---

<sup>136</sup> This is an exaggerated version of the result claimed by Weinberg, Nichols and Stich 2003. For a critical discussion of those results, see Nagel, J. 2012.

<sup>137</sup> This is an exaggerated version of the result claimed by Machery et al. 2004.

<sup>138</sup> Again, I'm assuming a realist construal of the subject matter of philosophical judgment. I'm also assuming that every philosophical proposition – or at least the ones believed by the Westerners and Easterners – is either true or false.

<sup>139</sup> To avoid spurious counterexamples, sensitivity is usually relativised to methods. I'll ignore this complication here.

<sup>140</sup> Another complication: applying a sensitivity condition to philosophical judgments involves evaluating counterfactuals with necessarily false antecedents, since (presumably) the propositions that are the contents of philosophical judgments are, if true, necessarily true. Thus, on a standard treatment of counterfactuals, sensitivity will be vacuously satisfied for all true philosophical judgments (Lewis 1973, 24-5.) If so, then even if sensitivity is a genuine condition on knowledge, the following argument will not be sound (cf. Nozick 1981, 322, 342ff). I ignore this complication here, granting for the sake of argument that our philosophical judgments are indeed insensitive. As I elaborate shortly, there are other reasons to reject (AI).

(AI) can also be used to target *evolutionarily* contingent judgments. Thus the sceptic might argue: “Take any one of your philosophical judgments, for example your judgment that we have special moral duties toward our children. The evolutionary evidence strongly suggests that humans’ judgments about their ethical duties toward their children are caused by natural selection rather than by the moral truths. This means that even if you didn’t have a special moral duty toward you children (holding the non-moral facts fixed), you would believe you did anyway!”

(AI) certainly enjoys a *prima facie* appeal. Indeed, it has proven a popular argument against moral realism (Blackburn 1985, 16-18; Williams 1985, 143, 151).<sup>141</sup> Nonetheless, many will reject (AI) on the grounds that sensitivity is an unpromising condition on knowledge. For sensitivity is well known to have wide-ranging sceptical consequences. Take my belief that I am not a handless-brain-in-a-vat. If I were a handless-brain-in-a-vat, I would still believe that I wasn’t a handless-brain-in-a-vat. My belief that I’m not a handless-brain-in-a-vat is insensitive to the truth of the belief. So, if sensitivity is required for knowledge, then I do not know that I am not a handless-brain-in-a-vat. If we accept the closure principle,<sup>142</sup> then it is also true that I do not know that I have hands – or very much else for that matter.

This is bad news for the defender of (AI). First, those philosophers who already reject sensitivity as a condition on knowledge will presumably be left cold by (AI). Second, genealogical sceptics themselves have reason to reject sensitivity, and thus

---

<sup>141</sup> Susan Hurley puts forward what she takes to be the best version of the evolutionary debunking argument against moral knowledge along sensitivity lines, though she ultimately endorses a non-sensitivity theory of epistemic justification (1989, ch. 14).

<sup>142</sup> Specifically, single-premise closure under known implication, viz. if I know *p*, and I know that *p* entails *q*, then I know *q*.

(AI), inasmuch as they themselves are not global sceptics.<sup>143</sup> Third, as an empirical matter, it's likely that judgments about sensitivity are themselves genealogically contingent in a way that makes them susceptible to (AI) – rendering (AI) self-defeating. Indeed, according to Weinberg, Nichols and Stich (2003), Westerners are more likely than South Asians to accept traditional, brain-in-a-vat style arguments. And it would be surprising if willingness amongst philosophers to accept sensitivity weren't in a part a function of various background factors.<sup>144</sup> If so, then (AI) impugns commitment to its own premises.

### 3.2 Explanatory Inertness

A second form of genealogical scepticism focuses on the supposed *explanatory inertness* of the (putative) truth of our philosophical judgments. According to this line of argument, our philosophical judgments are entirely explained by genealogies that make no mention of their truth; this in turn implies that we have no reason to believe that our philosophical judgments are correct. Take again our simplified case, in which you and your fellow Westerners share some philosophical judgment  $p$  while Easterners share the judgment  $\text{not-}p$ . We might think that the complete explanation of why you believe  $p$  rather than  $\text{not-}p$  is that you were born a Westerner rather than an Easterner. No mention here of the truth of  $p$ . Contrast the complete explanation of why I believe there is a computer in front of me. Presumably that explanation will draw some causal link between the fact that there is a computer in front of me, my sensory apparatus and my belief-forming mechanism. So unlike the explanation of your judgment that  $p$ , the explanation of

---

<sup>143</sup> Assuming they do not wish to reject closure.

<sup>144</sup> The PhilStudies Survey ([www.philpapers.org/surveys](http://www.philpapers.org/surveys)) has a wealth of information, for example, about the correlations between level of philosophical training and views on external world scepticism. The general tendency seems to be that the more philosophical training one has, the less likely one is to be an external world sceptic (22.6% of undergraduates say they lean toward or accept external world scepticism, while only 8.0% of graduates and 7.1% of faculty members/those with PhD's say the same). Within the group of faculty members/ PhD's, those who specialise in epistemology tend to have a higher than average acceptance of external world scepticism: 10.4%. Of those faculty members/PhD's who specialise in Continental philosophy, 11.4% lean toward or accept external world scepticism.

my judgment that there is a computer in front of me *does* mention the truth of that judgment. According to this line of argument, what is problematic is not genealogical contingency *simpliciter* – since my belief that there is a computer in front of me is also genealogically contingent – but rather judgments whose genealogies do not make mention of the *truth* of those judgments. We can lay out the argument as follows:

Argument from Explanatory Inertness (AEI)

- (7) Our philosophical judgments can be explained without mention of their (putative) truth
- (8) When a judgment can be explained without mention of its (putative) truth, then that judgment is unjustified
- (9) Therefore, our philosophical judgments are unjustified

Many have found (AEI) attractive, especially when applied to moral judgments. Thus here is Gilbert Harman, imagining the spontaneous moral judgment one would have encountering a group of children setting fire to a cat or a doctor cutting up a healthy patient in order to redistribute his organs:

[Suppose] you make a moral judgment immediately and without conscious reasoning, say, that the children are wrong to set the cat on fire or that the doctor would be wrong to cut up one healthy patient to save five dying patients. In order to explain your making the first of these judgments, it would be reasonable to assume, perhaps, that the children really are pouring gasoline on a cat and you are seeing them do it. But, in neither case is there any obvious reason to assume anything about “moral facts,” such as that is really is wrong to set the cat on fire or to cut up the patient...Indeed, an assumption about moral facts would seem to be totally irrelevant to the explanation of your making the judgment you make. It would seem that all we need assume is that you have certain more or less well articulated moral principles that are reflected in the judgments you make, based on your moral sensibility. It seems to be completely irrelevant to our explanation whether your intuitive immediate judgment is true or false (1977, 7).

Similarly, Street writes:<sup>145</sup>

...the best explanation of why human beings tend to make some evaluative judgements rather than others is not that these judgements constituted an awareness (however imperfect) of independent facts...but rather that the relevant evaluative tendencies forged links between our ancestors' circumstances and their responses which tended to promote reproductive success (2006, 114).

And Richard Joyce writes:

We have an empirically confirmed theory about where our moral judgments come from...This theory doesn't state or imply that they are true, it doesn't have as a background assumption that they are true...This amounts to the discovery that our moral beliefs are products of a process that is entirely independent of their truth, which forces the recognition that we have no grounds one way or the other for maintaining these beliefs. They *could* be true, but we have no reason for thinking so (2006, 211).

Since the (putative) truth of our moral judgments is inert in explaining those judgments, we lack reason to believe in their truth. As Harman writes: "Moral hypotheses do not help explain why people [make the moral judgments they do]. So ethics is problematic and nihilism must be taken seriously" (ibid., 11).

While (AEI) has proven particularly popular as an argument against moral realism, it might be easily expanded to other domains of philosophical judgment. The complete explanation of why you or I have the epistemic, metaphysical, aesthetic etc. judgments we do – in the terms of evolutionary psychology, or neuroscience, or sociology – will, we might suppose, make no mention of the *truth* of those judgments. So it turns out that epistemology, metaphysics, aesthetics and so on are problematic, and nihilism about those domains, as Harman says, is to be taken seriously.

---

<sup>145</sup> Although Street here gestures at a Harman-style explanatory inertness argument, this is not her main argument against moral realism. Rather, her main argument is motivated by the claim that there is no adequate explanation of how our moral beliefs, realistically construed, are reliable – that a 'massive coincidence' must be posited by moral realism. I discuss this style of argument in the next section.

What should we make of (AEI)? Both of its premises are contestable. (7) will come under pressure for any domain of philosophical judgment whose content is thought to be causally connected to judgments in that domain. Ethical naturalists, for example, identify moral properties (e.g. goodness) with causally efficacious non-moral properties (e.g. being conducive to human welfare). This identity secures, in turn, the causal efficacy of the moral properties. So, assuming ethical naturalism, it is possible that the complete explanation of ‘our’ moral judgments does in fact mention the moral truths.<sup>146</sup> As Nicholas Sturgeon (1988, 243-4) points out, moral facts seem to play a perfectly natural role in explanations of our beliefs: we believe Hitler was morally depraved because he was (cf. Wedgwood 2006; 2007, ch. 8). To claim, as Harman does, that the complete explanation of our moral judgments doesn’t actually feature moral facts might be thus thought to beg the question against the ethical naturalist.<sup>147</sup>

Take again my judgment that there is a computer in front of me, a supposedly unproblematic case of explanatory non-inertness. A sceptic about my computer judgment might argue as follows: “You believe there’s a computer in front of you just because of a particular configuration of your brain states that makes it *seem* to you that you are having the visual impression of a computer. This explanation makes no mention of the putative fact that there really *is* a computer in front of you. Therefore you are unjustified in believing that there is a computer in front of you.”<sup>148</sup> Harman might protest against the analogy. He might argue that this is

---

<sup>146</sup> Possible but not necessary, even assuming ethical naturalism, since one’s moral beliefs might be caused by forces apart from the causally efficacious moral truths – for example, by wishful thinking. Given that there is widespread disagreement about the moral truths, it cannot be that everyone’s moral beliefs are (non-deviantly) caused by the moral truths.

<sup>147</sup> Richard Joyce argues that careful reading suggests that Harman’s point is not that there are no moral facts, but more subtly, that if there is no naturalistic reduction of moral facts, then there are none (2006, 184-5). Thus Joyce’s own version of the explanatory inertness argument involves a putative refutation of ethical naturalism. At the risk of continuing the misrepresentation, I will take the standard line on Harman – that is, that he offers an argument for moral nihilism.

<sup>148</sup> Trenton Merricks (2001) and Ted Sider (forthcoming) argue for mereological nihilism on the grounds that it is the most parsimonious explanation of the evidence.

simply *not* the complete explanation of my judgment that there is a computer in front of me, for there is in fact a causal connection between there being a computer in front of me and my judgment that there is. But to assume, as Harman does, that a similar protest might not be launched against his scepticism about moral judgment is to beg the question against the ethical naturalist. For *if* there is a (non-deviant) causal connection between my moral judgments and the moral truths, as the ethical naturalist maintains is possible, then the complete explanation of my moral judgments will feature the truth of those judgments.

It's a point worth emphasising. Joyce (2006, chp. 6) argues that our mathematical judgments are not susceptible to his (AEI)-style argument because the complete explanation of why we have the mathematical beliefs we do is that evolution selected us to have *true* mathematical beliefs. After all, he points out, a creature who saw three predators enter the forest and two come out would be less likely to reproduce if it then inferred that the forest was clear of predators. It's undeniably useful to know that three less two is one. But a similar thing could be said of the moral truths. It's no less evolutionarily useful, after all, to know who is virtuous and who is vicious. Joyce's acceptance of the first argument but not the second presupposes that there are mathematical truths available to be tracked, but that there aren't any such moral truths. That might be so. But simply to *assume* so is to beg the question against the defender of causally efficacious moral truths.

We can generalise this dialectical observation. Inasmuch as one thinks that the domain of philosophical truths under sceptical pressure is causally active, one will also think it is explanatory alive, not inert – and thus that judgments about that domain might be robust against (AEI). For example, those who think that there exists a strong continuity between metaphysics or philosophy of mind and the natural sciences (and who are realists about the natural sciences) will likely think

that the truth of their judgments in those philosophical domains should be explained in ways not dissimilar to explanations of true scientific judgments.<sup>149</sup>

That all said, many philosophers are realists about domains of philosophical judgment that they think non-causal. For example, many are realists about ethics, epistemology and math without thinking that ethical, epistemic and mathematical objects are causally efficacious. Nonetheless, such philosophers will likely deny that judgments in these domains are undermined by (AEI). They might for instance object that the notion of explanation presupposed by (AEI) – that is, *causal* explanation – is unnecessarily narrow.<sup>150</sup> Thus one might take seriously the idea that particular philosophical judgments are best explained by their truth without thereby endorsing the view that such explanations are causal. Alternatively, one might argue along with Williamson that knowledge is explanatorily basic (2000); if our philosophical judgments are instances of knowledge, any putative explanation of them that does not explain why they are true fails to be an adequate explanation. Thus a purely psychological explanation might constitute an adequate explanation of why the brain-in-a-vat believes he has hands, but will not constitute an adequate explanation of why I, as an embodied creature, believe I have hands. Similarly, if our philosophical judgments are false, then an adequate explanation of them will not mention their truth; but if our philosophical judgments are items of *knowledge*, then an adequate explanation of them will. Even without supposing a causal relationship between the contents of our philosophical judgments and those judgments, we thus might have reason to deny (7).<sup>151</sup>

---

<sup>149</sup> For discussions of the extent to which metaphysics really is continuous with the natural sciences, see Ladyman and Ross 2007; Chalmers, Manley and Wasserman 2009; Dorr 2010.

<sup>150</sup> On the limitations of narrowly causal notions of explanation, see Nerlich 1979; Sober 1983; Woodward 2003, 220–22; Lipton 2004, 9–10.

<sup>151</sup> More generally, we might wonder why, since the truth can play an uncontroversial role in the explanation of factive propositional attitudes – e.g. *see that*, *happy that*, *regret that* – it cannot explain non-factive attitudes as well such as *believe that* and *judge that*?

Alternatively, the defender of philosophical judgments might accept (7) while rejecting (8), that the explanatory inertness of the truth of a set of judgments undermines their justification. Take for example my beliefs about the future – say, that I will be dining with Daniel tonight. The full explanation of that belief plausibly makes no mention of its truth; it will instead only feature the fact that I made a plan with Daniel, that I intend to keep it, that Daniel never misses a dinner plan, and so on. If explanatory inertness undermines justification, it seems that I cannot be justified in believing that I’m having dinner with Daniel tonight – or anything else, indeed, about the future.<sup>152</sup> So we again have some reason to doubt (8).

Whether (AEI) is sound thus turns on various issues in epistemology, as well as metaphysics: which properties and objects are causal, what constitutes a complete explanation, and whether explanatory inertness undermines judgment. Again, the upshot is two-fold. First, philosophers have at their disposal resources for rejecting (AEI). Second, (AEI) suffers from self-defeat. If our judgments about causality, explanation and justification are unjustified – as (9) implies – then it seems we would be unjustified in believing (AEI)’s own premises (cf. Pust 2001, Cuneo 2007).

### 3.3 Coincidence

A third sceptical argument appeals to the supposed difficulty – or indeed impossibility – of providing an explanation of how our philosophical judgments, given their particular genealogy, reliably track the truth. Though related in its focus on explanation, this argument is importantly distinct from that discussed in 3.2. There the argument presupposed that we could explain our philosophical

---

<sup>152</sup> An obvious way of amending (AEI) to avoid this implication is to shift the focus from explanatory inertness to an explanatory connection more generally. For my belief that I will have dinner with Daniel tonight and the fact that I will have dinner with Daniel tonight share a common cause, and thus enjoy an explanatory connection. See 3.3 for a discussion.

judgments without mention of their truth. The present argument instead relies on the distinct premise that there is no plausible explanation of how our philosophical judgments reliably get onto the truth: that the putative correlation between our philosophical judgments and the philosophical truths is a *coincidence*. One might deny the former while accepting the latter. For one might think that proximate explanations of particular philosophical judgments must include reference to the philosophical facts, but meanwhile deny that there exists a general explanation for how my philosophical judgments reliably track the truth.<sup>153</sup> If positing a coincidence between our judgments in a domain and the truth in a domain undermines our right to trust those judgments, then it seems that our philosophical judgments might be in peril.

The previous section's argument also relied on the premise that the explanatory inertness of the truth of a belief rendered that belief unjustified; as we saw, this had the unhappy implication that our beliefs about the future turn out to be unjustified. The current argument relies on the weaker premise that justification requires *some* explanatory connection between a belief and its truth. Since a third-factor or common cause explanation counts, this weaker premise does not undermine future beliefs (see n. 152).

Coincidence arguments are popular among genealogical sceptics.<sup>154</sup> For example, Ladyman and Ross claim that we have no reason to think that our metaphysical judgments are reliable, given that:

proficiency in inferring the large-scale and small-scale structure of our immediate environment, or any features of parts of the universe

---

<sup>153</sup> Only someone who endorses a non-causal view of the relevant domain of philosophical judgment would be likely to say something like this. For someone with a causal view will presumably claim both that individual judgments are best explained with reference to their truth *and* that the reliability of judgments in the domain as a whole can be explained causally. So I have in mind here someone who denies premise (7) of (AEI) but not because she thinks that the philosophical facts are causally efficacious.

<sup>154</sup> And, as I mentioned previously, debunkers of evolution like Thomas Nagel (see n. 123).

distant from our ancestral stomping grounds, was of no relevance to our ancestors' reproductive fitness (2007, 2).

Similarly, Street reasons that:

[T]he fact that there are any good scientific explanations of our evaluative judgements is a problem for the realist about value. It is a problem because realism must either view the causes described by these explanations as distorting, choosing the path that leads to normative skepticism or the claim of an incredible coincidence, or else it must enter into the game of scientific explanation, claiming that the truths it posits actually play a role in the explanation in question. The problem with this latter option, in turn, is that they don't. The best causal accounts of our evaluative judgements, whether Darwinian or otherwise, make no reference to the realist's independent evaluative truths (2006, 155).

And Hartry Field offers a similar argument against our claims to mathematical knowledge:

[O]ur belief in a theory should be undermined if the theory requires that it would be a huge coincidence if what we believed about its subject matter were correct. But mathematical theories, taken at face value, postulate mathematical objects that are mind-independent and bear no causal or spatiotemporal relations to us, or any other kinds of relations to us that would explain why our beliefs about them tend to be correct; it seems hard to give any account of our beliefs about these mathematical objects that doesn't make the correctness of the beliefs a huge coincidence (2005, 77; cf. Field 1989)

We might extend this style of argument to philosophical judgments in general:

#### Argument from Coincidence (AC)

(10) There is no plausible explanation of how our philosophical judgments reliably track the truth

(11) If there is no plausible explanation of how judgments in a domain track the truth in that domain, then those judgments are unjustified

(12) Our philosophical judgments are unjustified

This argument can be interpreted depending on how the ambiguity in ‘our’ philosophical judgments is resolved. When the philosophical judgments under attack are those shared by humans in general – our broad-brush epistemic, moral or metaphysical judgments – then the motivation for (10) will be that, *in light of the evolutionary origins of our philosophical judgments*, there is no plausible explanation of how they reliably track the truth. When the philosophical judgments under attack are those shared by a particular group – e.g. analytic philosophers, or Westerners – then the motivation for (10) will be that, *in light of the group-specific origins of our philosophical judgments*, there is no plausible explanation of how they reliably track the truth. In both cases, the conclusion is that our – whether that ‘our’ refers to humanity in general or some smaller subset – philosophical judgments are unjustified.

Again, those who favour causal accounts of particular domains of philosophical judgment will simply deny (10). For they will say there is a very good explanation of how our philosophical judgments reliably track the truth: they are causally related to the truth. But what of those philosophers who do not accept a causal account for a given domain of philosophical judgment? Must they accept that there is no plausible explanation for how judgments in that domain reliably track the truth, in light of their evolutionarily or culturally contingent origins?

Suppose that the movement of a certain commodities market in the Netherlands were isomorphic to the daily rainfall in the Amazon basin. What might explain the coordination? One possible explanation is causal: the exchange of the commodity affects the rainfall in the Amazon, or the Amazonian rainfall affects the exchange of the commodity, or there exists some third factor that causally influences both. Here is an alternative explanation, one that doesn’t posit a causal connection underlying the correlation: both Amazonian rainfall and the Dutch

commodity market are characterised by the same underlying mathematical structure. Thus their movements evince a striking isomorphism. Now consider a similar argument for the reliability of humanity's moral judgments: the moral truths are generally oriented around the principle of achieving social harmony, and evolution selected for creatures who had moral beliefs that were conducive to social harmony. Or one for the reliability of our epistemic judgments: the epistemic truths are generally oriented around truth-conduciveness, and evolution selected for creatures who are good at pursuing truth. Or one for the reliability of our mathematical and metaphysical judgments: the mathematical and metaphysical truths structure the way the world is, and evolution selected for creatures who were good at learning about the way the world is.

Analogous defences could be put forward against (AC) interpreted as an attack on the judgments of some subset of the human population. Let's suppose (AC) is run as an argument against the philosophical judgments of Westerners: given the cultural contingency of philosophical judgments, there is no plausible explanation for how Western philosophical judgments reliably track the truth while those of Easterners don't. But now suppose we respond: the moral truths are generally oriented around human harmony, the epistemic truths around truth-conduciveness, and so on....and Westerners have moral beliefs that are generally oriented around human harmony, epistemic beliefs that are oriented around truth-conduciveness, and so on. Thus the correlation is explained.

The genealogical sceptic might respond that, apart from being culturally chauvinistic,<sup>155</sup> this simply passes the explanatory buck. For what is now to be explained is the coincidence that the philosophical truths *are* such that they coincide with, say, what happens to be favoured by human evolution or Western culture. Take again the case of the Dutch commodity market and the Amazonian

---

<sup>155</sup> On the moral dimension of genealogically-motivated scepticism, see section 5.

rainfall. The sceptic might ask: but what explains the (putative) fact that these two phenomena share the same mathematical structure? Isn't that a coincidence? Indeed it is. But there is in general no prohibition against believing in coincidences. It's a coincidence that the musician Billy Bragg, whom I admire, and I were both born on 20th December. But it's not as if, absent a plausible explanation of this coincidence, I'm under pressure to stop believing in it. Explanation, after all, must come to an end somewhere.

Indeed, if the provision of explanation must eventually give out, then why not deny premise (II) of (AC)? Why not simply admit that the correlation between our philosophical judgments and the philosophical truths is *brute* – a 'massive coincidence', if you like, but one no more remarkable than the 'massive coincidence' that we're not brains-in-vats? If it's true that humans have generally reliable moral judgments, as we seem to, this is in part a matter of an absence of bad luck. And if philosophical training is in fact a help rather than a hindrance in getting onto the philosophical truths, as it seems to be, this is again a matter of luck. No doubt the proponent of (AC) – and genealogical scepticism more generally – is deeply uncomfortable with the idea that luck could play such a looming role in our capacity to know philosophical truths. But the extent to which luck plays a role in our ability to know is itself a philosophical question, one on which there is little epistemic consensus.<sup>156</sup> Thus, like the previous arguments we've examined, (AC) turns on a number of subtle questions of interest to epistemologists: about explanation, luck, justification and knowledge. Again, the upshot is two-fold.

First, those epistemologists who are already convinced of the ineradicable role that luck plays in our ability to have justified beliefs and knowledge – I think here mainly of epistemic externalists – will be unruffled by the revelation that they are

---

<sup>156</sup> See Essay Two, section 6 for a discussion of the epistemological debate about the role of luck in the acquisition of knowledge.

committed to a coincidence between our philosophical judgments and the philosophical truths. And they will be suspicious of any theorist's claim to have sidestepped such a coincidence. For explanation, as I've said, must come to an end somewhere. Those who claim to have removed the coincidence between a set of judgments and their truth – the cultural relativist, idealist or constructivist – will not be able to explain why we are the beneficiaries of the coincidence that makes knowledge in the relevant domain so very secure.

Second, (AC) is threatened with self-defeat. For if there exists no plausible explanation for how our philosophical judgments reliably get onto the truth, then this will presumably apply to our epistemic judgments as well, thus undermining commitments to (AC)'s own premises.<sup>157</sup> The proponent of (AC) might retort that the correlation between our epistemic judgments and their truth is merely brute, or that the correct explanation is simply that the epistemic truths are what they are, and our beliefs are what they are. But in either case, his protest is an instance of special pleading.

### 3.4 Evidential Improbability

A fourth sceptical strategy is to argue as follows: given the relevant genealogical evidence, it is improbable that our philosophical judgments are true. This argument again relies on the supposition that our philosophical judgments are not caused by the philosophical truths – a supposition rejected, as already discussed, by some philosophers for some domains of philosophical judgment. But let's suppose, *pace* such philosophers, that our philosophical judgments are indeed not

---

<sup>157</sup> Street (2006, n. 57) raises but does not adequately answer the worry that her debunking argument against moral realism is self-defeating. Her defence seems to be that, given her metanormative anti-realism, we have reason to accept her argument because we will in fact share the same or sufficiently similar epistemic judgments about its soundness. Thus perhaps Street's argument is most charitably read as an *ad hominem* attack against those who are antecedently committed to the epistemic premises of her argument. I discuss this in section 4. See especially n. 184.

caused by the philosophical truths. Instead, let us suppose that they are entirely caused by some relevant genealogy: a shared evolutionary or cultural history, philosophical training and so on. Now take some philosophical judgment of mine – for example, that Gettier’s Jones doesn’t know. Let’s suppose that I believe this because it was adaptive for my evolutionary ancestors to believe this, or because I was taught it by my culture, or in my philosophy class. How probable is it that this judgment is true, given that, wherever my belief came from, it has no causal relationship to the (putative) fact that Jones doesn’t know? It seems that the odds are no better than 50-50. When I then generalise the question to a whole domain of judgments – say, all my epistemological judgments – the chance gets very small indeed. Thus the probability that all or even most of my philosophical judgments are true is vanishingly low:

Argument from Probability on the Evidence (APE)

(I3) Conditional on the relevant genealogical evidence, it’s no more than 50% probable that any given one of one’s philosophical judgments is true

(I4) If it’s no more than 50% probable that one of one’s judgment is true conditional on the relevant genealogical evidence, then that judgment is unjustified

(I5) One’s philosophical judgments are unjustified

As noted, those who are committed to a causal connection obtaining between a given domain of philosophical judgment and truths in that domain will want to take issue with (I3). But what of those who are not so committed? Will they be persuaded by (APE)? They should, I think, dig their heels in at (I4). Suppose for instance that I *know* – as indeed I take myself to – that Jones doesn’t know.<sup>158</sup>

---

<sup>158</sup> Since knowledge requires reliability, and we are granting here that there is no causal relationship between the philosophical truths and my philosophical judgments, we might suppose

And let us suppose that everything I know is my evidence.<sup>159</sup> Then, how probable is it that my judgment is true, given all of my available evidence? Probability 1. If so, then the fact that, conditional only on the *genealogical* evidence, my judgment is no more than 50% likely to be true is irrelevant. Compare the following situation. Suppose that I know that Mary tells the truth exactly half the time. She tells me that Liz is in Hawaii. Conditional only on the fact that Mary told me so, that Liz is in Hawaii is 50% likely to be true. But conditional on my total evidence – which includes the fact that I just saw Liz in Oxford – the probability of Liz’s being in Hawaii is rather lower. When genealogical evidence isn’t our only evidence, the conditional probability of a particular belief’s being true on the genealogical evidence does not itself settle whether I should continue to hold onto that belief.

Of course, describing the evidential situation this way is to beg the question against the proponent of (APE). For the sceptic will deny that it *is* part of my evidence that Jones doesn’t know. My only evidence, he will say, is that I *judge* that Jones doesn’t know, and the genealogical fact that my judgment isn’t caused by its truth. But always to let the sceptic be the arbiter of what constitutes evidence is a dangerous policy. Williamson (2004a, 2000, 2007) argues that most forms of scepticism operate by narrowing the range of dialectically permissible evidence. Normally I take my total evidence to include facts about the external world. For example, I see Liz in the library, and use that as evidence to infer that she’s not in Hawaii, that she’s working on her dissertation, that she might fancy a drink later, and so on. But the sceptic narrows my evidence, granting only that it *seems* to me that Liz is in the library: how do I know that I am not hallucinating, or that Liz hasn’t a secret twin? He then challenges me to reason outward from my inner

---

that there is simply a brute correlation between my (species’, culture’s, academic community’s) epistemic judgments and the epistemic truths.

<sup>159</sup> For an argument for this identity, see Williamson 2000, sec. 9.7.

mental state of *seeming* to see Liz, to the (putative) external fact that Liz is indeed in the library.

We might say something similar about the dialectical situation involving the proponent of (APE). Normally, I include in my total evidence the fact that Jones doesn't know, using this evidence to infer that the justified-true-belief analysis of knowledge is false. But the proponent of (APE) insists that my only evidence is that it *seems* to me that Jones doesn't know. I am then challenged to reason outward from my mental state of judging that Jones doesn't know to the (putative) fact that Jones doesn't know. Williamson calls this sceptical manoeuvre the "psychologisation of evidence" (2004a, 118). Once our philosophical evidence is psychologised – transformed from the propositional content of our judgments to the bare fact that we have these particular judgments – we are in a quandary, for there is no clear route from the psychological fact to the philosophical fact.<sup>160</sup>

Our willingness to let our evidence be psychologised in the context of sceptical dialectic, argues Williamson, rests on two tacit principles, both of which he urges we reject. The first is that evidence is *operational*: that we're always in a position to know what our evidence is. This principle falls out of the common view that we must always be in a position to know what rationality demands of us; since epistemic rationality requires (at least in part) that we proportion our credences to the evidence, the operationalisability of rationality requires the luminosity of evidence. Take my seeing Liz in the library: let's call this the 'good case'. Call the 'bad case' the situation in which my counterpart hallucinates Liz in the library. Now suppose that my counterpart and I had different evidence. But then, on the assumption that evidence is operational, my bad case counterpart will be in a position to know that she is not in the good case, since she knows that her evidence is not what it would be if she were in the good case. But this is false:

---

<sup>160</sup> Save, of course, psychologising the subject matter of philosophy itself by reverting to the 'mere conceptual analysis' model. See section 2.1 for a discussion.

sceptics and non-sceptics alike agree that agents in sceptical scenarios aren't in a position to know that they are in sceptical scenarios, even if they paranoically believe they are. So it seems that there can be no difference in evidence between my sceptical counterpart and me; we each only know that it appears to us that Liz is in the library (Williamson 2000, ch. 8). But this argument fails, Williamson argues, because it is incompatible with the upshot of his anti-luminosity argument (Williamson 2000, ch. 4; Essay One). Our evidence – whether it be known propositions, beliefs, or qualia – isn't such that we are always in a position to know what it is. So there is nothing absurd about maintaining a difference in the evidence possessed by a normal agent and her sceptical scenario counterpart.

The second tacit principle that motivates us to psychologise our evidence is what Williamson calls *evidence neutrality* (2007, 210). Philosophy is a conversational business. In any fruitful conversation, certain agreed starting points are needed. Ideally, amongst these starting points should be a set of propositions that all parties can agree count as evidence; these bits of evidence should act as a neutral arbiter between the parties to the dispute. Thus

we might hope that whether a proposition constitutes evidence is *in principle* uncontentionally decidable, in the sense that a community of inquirers can always in principle achieve common knowledge as to whether any given proposition constitutes evidence for the inquiry. Call that idea *Evidence Neutrality* (ibid., 210).

The principle of evidence neutrality will naturally lead us to psychologise our evidence in order to converse with the (APE) sceptic. For he will deny that the fact that Jones knows is admissible as evidence. And it would be futile of me to insist otherwise, if our conversation is going to go on. Williamson does not deny that, for dialogue to be possible with the sceptic, a certain psychologisation of evidence is required. But he cautions against accepting the dialectical standard of evidence as *the* correct standard. One is sometimes justified, he argues, in rejecting the sceptic-approved basis of evidence as too “wilfully impoverished to constitute a

reasonable starting-point for evaluating the propositions in which one is interested” (2004a, 124).

Since genealogical sceptics of philosophy are not wholesale sceptics, they will need to explain why their psychologisation of the evidence is in principle different from the psychologisation of evidence employed by the traditional external world sceptic. For example, Joshua Alexander (2012, 104ff.) claims that for a philosopher to justifiably use her Gettier judgment as evidence, she must be able first to persuade an interlocutor who doesn't share that judgment of its truth. But a similar demand could be made for any premise whatsoever; if everything has to be defended to the satisfaction of the sceptic, then it will turn out that nothing is defensible. Why should we not include the contents of our philosophical judgments as evidence, but are permitted to include the putative facts that our judgments are such-and-such and have such-and-such genealogies? In other words, why should we let the sceptic about philosophical judgment set the appropriate standard of evidence, but not the wholesale sceptic?

The proponent of (APE) is on thin ice here. Anti-sceptical epistemologists like Williamson will not, consistently with their antecedent views, be sympathetic to (APE) in its resemblance to the wholesale sceptical argument. But to insist that (APE) is sound – and, in particular, that (I5) is true – is to insist on a philosophical claim. It's unclear how the proponent of (APE) expects his judgment that (I5) is true to pass the test set by (APE) itself. Conditional on the genealogical evidence, the probability that the sceptic's judgment that (I5) is true is itself true seems no better than 50%. If so, then (I5) implies that he has no reason to trust his judgment that (I5) is true. Lacking warrant himself, it seems that the (APE) sceptic has few resources to convince us of its soundness.

### 3.4.1. A detour: the Darwinian Dilemma, probabilised

Before moving on to my final version of genealogical scepticism – the argument from unreliability – I want to return briefly to my earlier discussion of Street’s evolutionary debunking argument against moral realism. An alternative way for Street to run her argument focuses not on the notions of coincidence and explanation, as she does, but instead on probabilities (cf. Copp 2008; Mogensen m.s.). Street might argue as follows: conditional on realism about moral truth and the evolutionary origins of our moral beliefs, it is highly improbable that our moral beliefs are generally true. By contrast, conditional on constructivism about moral truth and the evolutionary origins of our moral beliefs, it is highly probable that our moral beliefs are generally true. Given that our moral beliefs are generally true, we should reject realism in favour of constructivism.

How might the realist respond? One could respond, as Copp (2008) does, that to assume that the probability of our moral beliefs being true, conditional on realism and evolution, is near zero – for short,  $\Pr(T_m/(R\&E)) \approx 0$  – is to assume that realism is false. For suppose that realism were true: then our moral beliefs, *realistically* construed, would be generally true. Then it would also be true that the mind-independent moral truths were generally oriented around social harmony and flourishing. Given what we know about the workings of evolution – that it selects for creatures who are inclined toward social harmony – it seems not so likely that, conditional on realism and evolution, our moral beliefs are generally true. In other words, if realism is true, the objective chance that our moral beliefs are reliable is not anywhere near zero. So to assume that  $\Pr(T_m/(R\&E)) \approx 0$  is to beg the question against the realist.

In reply to Copp, Street (2008a) objects that this argument trivially begs the question against her, for it assumes certain substantive first-order moral views (i.e.

that social harmony is a good). There are a few things to say to this objection. The first is that some questions deserve to be begged, as is the case with the external world sceptic (or, even more clearly, the sceptic about reason); if the sceptic-approved evidential base is too narrow, the anti-sceptic is justified in refusing to take it as her starting point. This might be one of those cases. The second point is this: Street's constructivism has various substantive first-order moral entailments, e.g. "Dani ought to drink tea iff he would believe he ought to drink tea in reflective equilibrium". So in claiming that the conditional probability of our having reliable moral beliefs on constructivism plus evolution is close to one – for short,  $\Pr(T_m/(C\&E)) \approx 1$  – we might think Street too assumes certain substantive first-order moral views. No doubt Street will reply that her constructivism doesn't involve any assumptions about the substantive *content* of morality, but is instead just a metaethical claim about what *makes* the moral truths true (cf. Street 2006, n. 56). It's unclear whether such a distinction can be coherently maintained.<sup>161</sup> But in any case, and thirdly, Street's charge of question-begging seems off the mark. For Copp *doesn't* assume that our moral beliefs are reliable. He rather only claims that *if* our moral beliefs are reliable, then the objective chance that they are reliable conditional on realism and evolution will hardly be near zero. Thus Street must offer a non-question begging reason for why realists should accept the crucial premise of her argument that  $\Pr(T_m/(R\&E)) \approx 0$ .<sup>162</sup>

A more general point can be made here. Probabilistic arguments in favour of anti-realist over realist treatments of philosophical judgments tend to beg the question against realism. For, without assuming that our substantive philosophical judgments are false, it is hard to see how the anti-realist can motivate his claim that the chances of our philosophical judgments being true, conditional on realism about those judgments and their respective genealogies, is near zero.

---

<sup>161</sup> There is a live debate about whether the first-order/meta distinction can be coherently drawn in ethics at all. See e.g. Dworkin 1996; Blackburn 1998; Bloomfield 2009; Kramer 2009.

<sup>162</sup> Mogensen (m.s.) argues that this premise can be defended by appeal to a principle of indifference, but then offers a compelling case that we should reject that appeal.

### 3.5 Unreliability

I have saved what I take to be the most promising argument for genealogical scepticism for last. The argument deploys the widely-accepted *safety* condition on knowledge, according to which S knows *p* only if S could not have easily falsely believed *p* using a sufficiently similar method to the one she actually uses to believe *p* (Sosa 1999). Put more simply, safety demands that one not have ‘nearby false beliefs’: one’s belief must be *reliable*. Suppose Ahmed is watching a coin being flipped. Because Ahmed prefers heads to tails – say, he stands to win some money if the coin comes up at heads – Ahmed forms the belief that the coin will come up heads. And indeed it will. But Ahmed doesn’t of course know that the coin will come up heads. The safety condition gives a plausible diagnosis of why not: there is a nearby world in which Ahmed wishfully believes the coin will come up heads but in which it will not. Thus Ahmed has a nearby false belief that the coin will come up heads, rendering his actual true belief unsafe.

Unlike the sensitivity condition on knowledge, the safety condition does not swiftly generate the result that we know nothing at all. Return to the brain-in-a-vat scenario. My belief that I’m not a brain-in-a-vat is insensitive: if I were a brain-in-a-vat, I would still believe that I wasn’t. So according to the sensitivity condition on knowledge, I’m not in a position to know that I’m not a brain-in-a-vat. But my belief that I’m not a brain-in-a-vat is – on the assumption that I’m in fact embodied, and that there’s no nearby threat of being envatted – safe. So safety has the benefit of capturing the intuitive importance of reliable truth-tracking for knowledge but, because that demand is confined to an agent’s modal neighbourhood, does not generate sceptical worries.

Safety is generally considered an *externalist* condition on justification. According to epistemic *internalism*, justification supervenes entirely on states that are suitably ‘internal’ to agents. How this notion of ‘internal’ is unpacked generates two varieties of internalism. According to ‘access’ internalism, justification supervenes solely on states to which agents have privileged first-personal access, i.e. some subset of the conscious mental states (Chisholm 1977; Bonjour 1985).<sup>163</sup> According to ‘mental state’ internalism, justification supervenes solely on a certain cluster of non-factive mental states, whether or not agents enjoy privileged access to those states (Conee and Feldman 2001; cf. Wedgwood 2002<sup>164</sup>). Externalism is the denial of the claim that justification supervenes entirely on non-factive mental states. Safety is an externalist condition on justification because it is not a mental state.

But safety considerations can be deployed by internalists too. For most internalists will agree that strong evidence that one’s belief is unsafe defeats the justificatory status of one’s belief. That is, even if one’s belief was previously (internalistically) justified, acquiring evidence that it is unsafe renders it (internalistically) unjustified. Suppose Charlotte sees a red book on the laboratory table and forms the true belief that there is a red book on the table. Then her chemistry teacher – whom Charlotte has every reason to believe is trustworthy and reliable – informs her that there is a red light shining on the table, a light that makes non-red objects appear red. This constitutes evidence that Charlotte’s belief is unsafe; Charlotte now has reason to believe that her visual inspection of the book on the table could have easily led her to have a false belief. Most internalists will want to say that, so long as Charlotte has no reason to mistrust her teacher, and has no reasons independent of her visual inspection for believing that there is a red book on the

---

<sup>163</sup> If Williamson’s anti-luminosity argument is sound, then access internalism implies nihilism about epistemic justification, since there is no suitable supervenience base for justification. See Essay Two for a discussion of the anti-luminosity argument and its bearing on the internalist/externalist distinction.

<sup>164</sup> Wedgwood advocates a mental state internalism about rationality rather than justification.

table, her belief is rendered unjustified by the new evidence she has acquired. So while internalists won't endorse a safety condition on knowledge, they will most likely endorse what we might call an *internalist safety* condition, according to which strong evidence of unsafety defeats justification.

What will an externalist want to say about this sort of case? First we must stipulate whether it is in fact true that there is a red light shining on the table. If there is, then the teacher's testimony is truthful. If on the other hand there is no such light, then the teacher's testimony is misleading. (We need not suppose the teacher is being *intentionally* misleading; he might have been misled himself.) Suppose the teacher's testimony is truthful: there really is a red light shining on the table. Then it is the case that Charlotte's belief is unsafe, and according to the externalist, unjustified. Importantly, this was true *before* Charlotte received the teacher's testimony. The receipt of the testimony simply allows Charlotte to come to know that her belief is, and indeed has always been, unjustified.

Now suppose that there is in fact no red light, and thus that the teacher's testimony is misleading. What will the externalist say about this? Here things are trickier. First, we can distinguish between those externalists who do and do not countenance the possibility of *defeat*, the loss of justification through the mere acquisition of new evidence.<sup>165</sup> Defeatist externalists will say that, although the teacher's testimony is misleading, it nonetheless destroys Charlotte's justification (Goldman 1986, 62-3, 111-2; Alston 1988b, 238-9; Nozick 1981, 196). Thus the defeatist externalist gives the same verdict as our internalist.

But what of those externalists who reject defeat? Here we can further distinguish between *dogmatic* anti-defeatists and *akratic* anti-defeatists. Dogmatic anti-

---

<sup>165</sup> I say 'mere' to distinguish genuine defeat cases from cases in which an agent loses justification because the acquisition of new evidence results in the agent's ceasing to believe or changing her belief-forming method or basis for belief.

defeatists think that one can retain the epistemic status of one's belief – as justified, or warranted or knowledge – so long as one dismisses the misleading evidence. Thus dogmatists will say that Charlotte can dismiss her teacher's misleading testimony, thereby retaining her justified belief that there is a red book on the table.<sup>166</sup> By contrast, akratic anti-defeatists think, in such cases, that one should not dismiss the misleading evidence; if one believes anything about one's first-order belief, one should believe that it lacks the relevant epistemic status. But compatible with that, the akratic anti-defeatist will say, is maintaining a justified (or warranted, or knowledgable, or rational) first-order belief. Thus Charlotte can justifiably believe that there is a red book on the table, while also justifiably believing that *that* belief is unjustified.

The safety-based argument I'm going to formulate on behalf of the genealogical sceptic will aim to be acceptable to both internalists and defeat-friendly externalists. I'll return to the discussion of anti-defeatists – of both the dogmatic and akratic variety – later. But first, one more preliminary. The propositions under discussion in this paper – namely, philosophical propositions – are plausibly *necessary*; if Gettier's Jones doesn't know, it is *necessarily* the case that he doesn't know. Beliefs about necessary propositions will trivially satisfy Sosa's safety condition.<sup>167</sup> But a related condition can be constructed for beliefs about necessary propositions:

SAFETY<sub>n</sub>: S's belief in the necessary proposition *p* is safe<sub>n</sub> iff S could not have easily believed not-*p* using a sufficiently similar method she uses to believe *p*.

---

<sup>166</sup> Some dogmatists, for example Plantinga and Thomas Nagel, think that only certain kinds of judgments are immune from defeat. I discuss this in more detail shortly.

<sup>167</sup> A similar problem arises for appeal to sensitivity conditions in sceptical arguments against philosophical judgment. See n. 140.

Safety<sub>n</sub>, like its safety cousin, seems intuitively right. Suppose I am very poor at math. I want to know the sum of 236 and 582. After struggling to add the sums in my head, I take a guess and come to believe it's 818. I now believe a necessary truth. My belief is thus trivially safe, since there are no nearby (or indeed any) worlds in which 236 plus 582 doesn't equal 818. But my belief is not safe<sub>n</sub>. For the method I used to arrive at my true belief – struggling a bit and then taking a guess – easily leads me to believe that the sum of 236 and 582 is *not* 818. Or take my Gettier judgment. Let's suppose that I believe that Jones doesn't know simply because my professor told me so; had my professor told me that Jones did know, I would have believed that. Again, it seems that I don't know, and the reason seems to be because my judgment fails safety<sub>n</sub>: in a very nearby world, I use the method of trusting whatever my professor says to come to believe, falsely, that Jones does know.<sup>168</sup>

Let's use this notion to construct a final argument for genealogical scepticism:

#### Argument from Unreliability (AU)

- (16) The genealogy of our philosophical judgments constitutes strong, undefeated evidence that those judgments are unsafe<sub>n</sub>
- (17) Whenever we have strong, undefeated evidence that one of our judgments is unsafe<sub>n</sub>, we ought to abandon it
- (18) We ought to abandon our philosophical judgments

Again, this argument can be unpacked depending on what sort of specific genealogical debunking claim is at issue. With experimental debunking, 'our' refers to the philosophical judgments shared by some group. So the experimental

---

<sup>168</sup> When making judgments about safety (and reliability more generally), how we fill out the cases matter. Say my professor turns out to be omniscient. Then my Gettier judgment seems very secure indeed, hardly unsafe<sub>n</sub>. What if my professor isn't omniscient, but is more mundanely extremely reliable with philosophical judgments and is also always sincere? Maybe in this case my Gettier judgment again turns out to be safe<sub>n</sub>. Here I am assuming that my professor is either a not very reliable judge or not reliably sincere.

philosopher can argue that his data – that philosophical judgments systematically vary with culture, gender, socioeconomic status, philosophical training and so on – gives you (say, a privileged white male professional philosopher) decisive reason to believe that your philosophical judgments are unsafe<sub>n</sub>. This is because the experimental data suggest that, using the method you actually use to arrive at your philosophical judgments, you could have easily believed the opposite of those judgments – if you were poor, non-white, female or philosophically untrained. Since the experimental data give you decisive reason to believe that your philosophical judgments are unreliable, you ought to abandon those judgments.

The evolutionary debunking version of (AU) will run similarly. The evolutionary data – that we have the philosophical judgments we do because being disposed to have such judgments was adaptive for our ancestors – provides us with strong, undefeated evidence that our philosophical judgments are unsafe<sub>n</sub>. Why? Because our evolutionary history could have very easily gone differently, and had it done so, our method of philosophical judgment would have given rise to contrary philosophical judgments. As Darwin himself speculated:

In the same manner as various animals have some sense of beauty, though they admire widely different objects, so they might have a sense of right and wrong, though led by it to follow widely different lines of conduct. If, for instance, to take an extreme case, men were reared under precisely the same conditions as hive-bees, there can hardly be a doubt that our unmarried females would, like the worker-bees, think it a sacred duty to kill their brothers, and mothers would strive to kill their fertile daughters; and no one would think of interfering. Nevertheless, the bee, or any other social animal, would gain in our supposed case, as it appears to me, some feeling of right or wrong, or a conscience (1879/2004, 121-122).

If it's true that we could have easily had opposing moral or other philosophical judgments, then it seems that the method we use to arrive at our actual philosophical judgments – roughly, the method of believing in accordance with our evolutionarily-selected dispositions to judge – could easily lead us into falsity.

Recognising this, the evolutionary debunker might argue, we should abandon our philosophical judgments.<sup>169</sup>

Interestingly, a version of (AU) can also be deployed to support the conciliatory verdict on the question of peer disagreement – roughly, the verdict that disagreement with an epistemic peer defeats one’s justification (Feldman 2006; Elga 2007; Christensen 2007a, 2010, 2011). Suppose you and your best friend Hussein always go to lunch and split the bill. You have equally reliable track records when it comes to doing the mental math required to figure out how much you each owe. On this occasion, you believe that you each owe X amount and Hussein believes that you each owe Y amount (where X doesn’t equal Y). Many think that, in such a case, you ought to abandon your belief that you each owe X amount, and only re-instate it after you have double-checked or achieved some independent confirmation that you were right.<sup>170</sup>

(AU) gives us a promising way of vindicating this thought. For we might think that the fact that Hussein disagrees with you constitutes strong undefeated evidence that your judgment (that you each owe X amount) is unreliable. For Hussein’s disagreement suggests that you could have easily, using the method of mental math, arrived at the belief that you each owe not-X. In the face of this decisive reason to believe your judgment is unsafe<sub>n</sub>, you ought to (at least temporarily) abandon it.

---

<sup>169</sup> It’s worth noting that there is nothing about (AU) that requires *systematic* variation in philosophical judgments. Suppose that there was simply random disagreement about philosophical questions that did not track any other (cultural, gender, training, and so on) vectors. This would plausibly still constitute *prima facie* evidence of the unreliability of these judgments. But the systematic nature of the variation does have a certain dialectical power, since it forces philosophers effectively to assert the superiority of their methods *over* the methods of non-philosophers, or non-Westerners, etc. I discuss the significance of the moral dimension of genealogical scepticism in section 4.

<sup>170</sup> This case I borrow from Christensen 2011.

Before discussing how the defender of philosophical judgment might respond to (AU), I want to point out that the present argument signals an interesting shift in the flow of discussion. My suspicion is that what people find intuitively compelling about genealogical scepticism doesn't have much to do with considerations of safety. Instead I suspect that when considering evolutionary or experimental debunking arguments, people are struck by the (putative) revelation that what is 'really doing the work' behind our beliefs are forces orthogonal to the truth of those beliefs – forces of evolution, culture and so on. This then sends one into sceptical free-fall: *If our beliefs have nothing to do with the truth, how can we trust them?* This thought can be cashed out in various ways: in terms of explanatory inertness, or coincidence, or probabilities. As I've discussed, I don't think those arguments should be convincing to many philosophers. (That said, they might still be sound.) One dialectical weakness they share is the assumption that there is no causal connection between our philosophical judgments and the philosophical truths – an assumption denied by many for many domains of philosophical judgment.

By contrast, (AU) doesn't suppose that no causal connection exists between our philosophical judgments and the philosophical truths. Rather, it suggests, more weakly, that even if there is some sort of causal connection between the philosophical truths and our philosophical judgments, that connection is insufficiently reliable to produce knowledge. In a sense it's a less thrilling point to make; it doesn't inspire the same sort of sceptical vertigo as the earlier arguments. But this, I submit, is a feature rather than a bug. Defending our philosophical judgments against (AU) is a subtler and more interesting business. I'll now turn to how that might be done.

Recall our branching taxonomy of epistemologists. First we have the contrast between internalists and externalists. Within externalists we have defeatists and

anti-defeatists. And finally we have both dogmatic and akratic anti-defeatists. So we have: internalists, defeatist externalists, dogmatic anti-defeatist externalists and akratic anti-defeatist externalists. I'm going to discuss how each kind of epistemologist might respond to (AU) in turn. Because they handle misleading higher-order evidence in the same way, I'll group together internalists and defeatist externalists.

### 3.5.1. Internalists and defeatist externalists

Premise (I6) says that the genealogy of our philosophical judgments constitutes strong undefeated evidence that those judgments are unsafe<sub>n</sub>. If that's true, then internalists and defeatist externalists alike will conclude that we ought to abandon those judgments.<sup>171</sup> But why think (I6) true? Why think that the relevant genealogical evidence *is* evidence that our philosophical judgments are unsafe<sub>n</sub>?

Take the putative variation in Gettier judgments with extent of philosophical training. To simplify matters, suppose that those with philosophical training share the judgment that Jones doesn't know, and that those without philosophical training judge that Jones does know. The proponent of (AU) thinks that this constitutes evidence that our (that is, we philosophers') Gettier judgments are unsafe<sub>n</sub>. But this is only true if the methods used by philosophers and non-philosophers to arrive at their respective Gettier judgments are indeed 'sufficiently similar'. One might think instead that two quite dissimilar methods are in play here – the expert method employed by philosophers, and the lay method used by the philosophically untrained (Bealer 1998; Devitt 2006, 2011; Kauppinen 2007;

---

<sup>171</sup> This is an oversimplification. Wedgwood for example is an internalist who endorses akrasia about (both deductive and non-deductive) inference (2011). That is, he thinks that certain beliefs formed on the basis of competent inference, unlike perceptual beliefs, are not susceptible to defeat. Thus Wedgwood will think that at least some of our 'philosophical judgments', if the result of competent inference, will be immune from defeat. For the sake of simplicity I am going to bracket this sort of internalist view.

Sosa 2007; Williamson 2007, 191-2; Liao 2008; Pinillos et al. 2011)<sup>172</sup> Similar things could be said in response to other instances of (AU). For example, one might think that the method we humans use to arrive at our moral judgments is rather dissimilar from the method that would be used by Darwin's imaginary Bee People. Our method, we might say, is one that is sensitive to genuine moral reasons, while the method used by the hypothetical bee people lacks such sensitivity. And so on. If so, then (16) is false: the genealogical evidence is *not* evidence that our philosophical judgments are unreliable.

Whenever we speak of methods of belief-formation in accounts of epistemic reliability we come up against the vexing *generality problem* (Goldman 1979; Conee and Feldman 1998). For how are we to know how to type the relevant methods when evaluating the safety of a given belief? In general, a very coarse-grained typing of 'sufficiently similar method' will give us the absurd conclusion that practically no belief is reliable, while a maximally fine-grained typing will give us the equally absurd conclusion that every belief is reliable.<sup>173</sup> According to Williamson (2000), we should not think of this as a damning problem. Or rather, we should only think it damning if we were hoping that safety would give us an entirely independent grasp on whether particular beliefs constitute knowledge. Instead, we should think of our judgments about knowledge and methods as mutually informing. Whether or not the generality problem is a problem for safety (and other reliabilist) accounts of knowledge, it does pose a dialectical challenge

---

<sup>172</sup> Alvin Plantinga (1995) gives a similar response to genealogical attacks on the rationality of religious exclusivism. He argues that it might be the case that the Christian is using a superior method in arriving at her Christian belief, and that the non-Christian "has *made a mistake*, or *has a blind spot*, or hasn't been wholly attentive, or hasn't received some grace she has, or is in some way epistemically less fortunate" (205).

<sup>173</sup> A similar problem arises for the question of which worlds count as 'nearby' in modal conditions such as safety and safety<sub>n</sub>. This would be an alternative way for defenders of philosophical judgment to go, i.e. arguing that the worlds in which they would have mistaken philosophical judgments are *too far away* to undermine the reliability of their actual judgments. I find this kind of defence more plausible against evolutionary debunking arguments than experimental debunking arguments. That is, it's more plausible that the world in which I am, say, a Bee Person is too distant to undermine my moral judgment, than it is to say that the world in which I'm, say, Chinese is similarly too distant.

here. For the philosopher will take her philosophical judgments to constitute knowledge, and will on that basis judge that the methods in question are sufficiently dissimilar not to undermine that knowledge. Meanwhile the sceptic will deny that these philosophical judgments do constitute knowledge, and so will insist that the relevant methods are to be typed more coarsely. The generality problem means that there is no dialectically neutral way of settling this question. So again, the sceptic and defender of philosophical judgment will find themselves begging the question against each other. Given that the sceptic is the one who hopes to convince his opponent, this seems worse news for him.

### 3.5.2. Dogmatic anti-defeatism

According to Thomas Nagel, certain beliefs – well-formed beliefs about moral, logical and mathematical truths – are simply immune from defeat (1996). He thus rejects (17): at least some of our genealogically contingent philosophical judgments are unscathed by (AU). Plantinga meanwhile argues that certain judgments – ones that are ‘properly basic’ – are what he calls “intrinsic defeaters” (1986, 311), meaning that their contents defeat any potential defeaters. Thus, assuming that our philosophical judgments are properly basic,<sup>174</sup> Plantinga will reject (16): the genealogical evidence does not constitute *undefeated* evidence that our philosophical judgments are unreliable. Maria Aarnio-Lasonen (2011) puts forward a more general anti-defeat case, arguing that we can always retain knowledgable belief by acting dogmatically in the face of misleading evidence.<sup>175,176</sup>

---

<sup>174</sup> Plantinga believes that many of our philosophical judgments are properly basic – for example, our moral judgments.

<sup>175</sup> Though in her (forthcoming), Aarnio-Lasonen seems more sympathetic to an akratic form of anti-defeatism.

<sup>176</sup> It’s important to note that while Aarnio-Lasonen thinks it is possible to retain knowledge in the face of misleading evidence, she thinks that doing so is often *unreasonable* and thus epistemically criticisable. So it’s not clear how much comfort her view offers to the philosopher who wishes to hold onto her philosophical judgments in the face of (possibly misleading) genealogical evidence.

For the sake of simplicity I'm going to focus on Plantinga's form of dogmatic anti-defeatism. Suppose that I'm confronted with evidence that my judgments about Gettier cases are unreliable, for it turns out that those judgments are culturally contingent. According to Plantinga, if my Gettier judgments are in fact properly formed – and thus the genealogical evidence misleading – then I can use the contents of those very judgments to dismiss the misleading evidence. Thus I can reason as follows: *The genealogical evidence suggests that my Gettier judgments are unreliable. But in Gettier cases the subjects don't know. So my Gettier judgments are reliable, and the genealogical evidence is misleading.*

The sceptic will protest that this is a form of *bootstrapping*, using a method to confirm its own reliability without any independent confirmation. Thus Robert Cummins complains that

we have no access to the workings of intuition that is *independent* of intuition itself; thus, intuition cannot be calibrated against errors in the way that a scientific device can (1998, 116-7, italics added).

Jonathan Weinberg makes a similar point in his discussion of philosophical expertise:

Trained expert judgments, for example, like those of chess grand masters or medical diagnosticians, lie outside the range of my critique; so too do our judgments in most ordinary cases that some particular object or event falls under a particular concept. I am not attacking such intuitions as those not because they are immune to worries about hopefulness, but because by and large they are in fact hopeful. Both expert judgments and ordinary categorizations usually possess a great deal of *external corroboration* and internal coherence (2007, 334-5, italics added).

Weinberg's implication is that, unlike the doctor or chess master, the philosopher lacks recourse to suitably *independent* corroboration of the reliability of her judgments. Of course much hangs on what we consider an 'independent' means of corroboration. When I use my sensory apparatus – say, in scientific

experimentation – to confirm the reliability of the human sensory apparatus, am I thereby illicitly bootstrapping my way into ‘easy knowledge’? Presumably not. Compare the case of my Gettier judgment. When I consider the Gettier case, I find myself strongly disposed to judge that Jones doesn’t know. Suppose, after forming my Gettier judgment, I then think abstractly about the nature of knowledge, and conclude that it is incompatible with certain forms of luck. Then I return to the Gettier case and realise that it involves a certain form of luck. Might not this serve as independent corroboration of my initial judgment?

For the sake of argument let’s grant the sceptic that there is no suitably ‘independent’ means of corroborating the reliability of our philosophical judgments. If they are to be corroborated at all, it will be through a form of bootstrapping. Is this worrying? Whether bootstrapping is illicit or not is a matter of epistemological controversy. Jonathan Vogel (2000), Stewart Cohen (2002) and Roger White (2006) agree that bootstrapping – and the ‘easy knowledge’ it produces – are epistemically unkosher. Meanwhile Plantinga (1995) defends the use of bootstrapping in cases of religious and moral disagreement, and James van Cleve (2003) argues that if bootstrapping is illicit, the only alternative is scepticism. Without getting into the details of the relevant arguments, I want to offer a case that puts pressure on the view that bootstrapping is always epistemically illicit:

**SURPRISE HOMOPHOBE:** You have known your friend Robert for a few years now, and you believe his judgment to be as good as yours. For you agree on most things, and when you occasionally don’t, one is always able to change the mind of the other through reasoned conversation. One day you’re shocked to discover that Robert thinks that homosexuality is immoral. You discuss this for some time and realise your disagreement is bedrock. Robert simply has the strong, intuitive judgment that homosexuality is immoral. But on all other issues you continue to agree.

What should your response be? Should you conclude that your judgment about homosexuality is unreliable, and therefore that you should abandon it? I think not. You should rather conclude that Robert's judgment on this question is impaired, perhaps because he has grown up in a bigoted home or is suffering from repression and self-hatred. Crucially, the only reason you have for concluding that Robert's judgment is impaired – and thus that the disagreement is misleading evidence of your own unreliability – is the fact that homosexuality is *not* immoral.

Arguments about (non-moral) philosophical issues are generally far less emotionally charged than this example.<sup>177</sup> But if bootstrapping is acceptable in some cases, we should ask to know why it is not acceptable in others. The (AU)-sceptic cannot simply assume the illegitimacy of bootstrapping, or the falsity of dogmatic anti-defeatism – or at least, as I will argue shortly, not without risking self-defeat.

### 3.5.3 Akratic anti-defeatism

A final response to (AU) comes from the akratic anti-defeatist. Unlike the internalist, defeatist externalist or dogmatic anti-defeatist, the akratic anti-defeatist denies (17). That is, she thinks that it can be rationally permissible at the same time to (a) believe  $p$ ,<sup>178</sup> and (b) believe that one's belief that  $p$  is unjustified. While many epistemologists assume or argue for the falsity of the akratic position (Feldman 2005; Kolodny 2005; Christensen 2007a, 2007b, 2010, 2011; Elga 2007; Smithies 2012; Greco forthcoming; Horowitz forthcoming; Titelbaum m.s.), there is a growing pro-akrasia contingent.

For example, Allen Coates (2012) and Brian Weatherson (m.s.) offer general

---

<sup>177</sup> In section 5 I discuss why the moral dimension of this case is so important to its (I hope) persuasiveness.

<sup>178</sup> Or if we're speaking in terms of credences rather than outright belief, to maintain the same credence in  $p$ .

arguments for the akratic conclusion that, in cases of misleading higher-order evidence, one can justifiably believe  $p$  and justifiably believe that one's belief that  $p$  is unjustified. In such cases, Weatherson claims, "each conjunct is well-supported by the evidence" (ibid., 15) and thus rationally permissible to believe. And Coates claims that when higher-order evidence is misleading "we can rationally judge that our belief is irrational even though it is in fact rational, and so we can rationally be akratic" (2012, 122). Williamson (2011) argues that a long competent deduction is a case in which epistemic akrasia is rational, since even if one has achieved knowledge through the deduction, one is still in a position to justifiably believe that one has made a mistake in one's reasoning somewhere. Similarly, Wedgwood argues that a competent inference produces knowledge even when an agent rationally believes that it hasn't, or rationally doubts that it has.<sup>179</sup> Thus Wedgwood writes:

a perfectly rational thinker would continue to draw the inference even if she had (misleading) evidence that she was reasoning incompetently, and even if she entertained serious doubts about whether or not she really was perfectly rational (2011, 21).

Williamson (2011) further argues that even when our perceptual evidence is strong enough to allow us to know, that evidence often makes it improbable that we *do* know. This is because we are not always, according to Williamson, in a position to know what our evidence is; thus our evidence can be massively misleading about the quality of our epistemic situation. To deny the rational permissibility of epistemic akrasia is, he argues, tantamount to scepticism about perceptual knowledge.

---

<sup>179</sup> Wedgwood is an internalist who accepts the possibility of defeat for perceptual knowledge; he thinks that certain cases of competent inferences are special and thus license akratic treatment (see n. 171).

One might hope to extend a broadly pro-akrasia view to genealogical evidence about philosophical judgment.<sup>180</sup> Thus one might think that we are justified in believing, in light of the genealogical evidence, that our philosophical judgments are unjustified – but nonetheless think that, so long as those judgments *are* justified, we are permitted to maintain them. No doubt there is something uncomfortable about asserting and theorising on the basis of judgments one takes oneself probably not to know. And it's an interesting question whether, as a psychological matter, one could really go on philosophising in this way. But for those who think that epistemic akrasia is sometimes rational, this might be a final way of resisting the pull of (AU)-scepticism.

\*

Whether (AU) is sound turns on various complex issues in epistemology: about method individuation, the generality problem, varieties of internalism and externalism, dogmatism, bootstrapping, akrasia and defeat. Again, this has a two-fold upshot. First, philosophers have at their disposal various philosophical resources for resisting (AU); and indeed many philosophers (especially those sympathetic to epistemic externalism) will antecedently be committed to views that clash with (AU). Second, the proponent of (AU), like the other kinds of genealogical sceptic I've discussed, is at risk of self-defeat. People's judgments about (16) and (17) systematically vary; for example, willingness to accept or deny (17) is likely strongly correlated with whether one has studied philosophy at Oxford. Does this not constitute strong, undefeated evidence that the sceptic's judgment that (16) and (17) are true is itself unsafe?

---

<sup>180</sup> Whether one can extend the particular arguments marshalled in favour of epistemic akrasia by Coates, Weatherson, Wedgwood and Williamson will turn on various questions about how we should think of what is going on in the formation of philosophical judgment – whether, for example, all philosophical judgments should be classified under Wedgwood's heading of indefeasible inference, or whether philosophical judgments should be thought of as evidentially-based (Coates and Weatherson). Note that Williamson's 'improbable knowing' argument doesn't easily extend to philosophical judgments since it crucially involves contingent truths and the notion of a margin-for-error.

The sceptic will perhaps protest that his judgments that (I6) and (I7) are true are not susceptible to (AU). For he might claim that these judgments are the result of an *expert* method. Or the sceptic might argue that while the variation in judgments about (I6) and (I7) provides *prima facie* evidence that his judgments are unsafe<sub>n</sub>, that evidence is defeated by the truth of (I6) and (I7). Or finally the sceptic might concede that his judgments are likely unsafe<sub>n</sub>, but continue theorising on the basis of those judgments in the hope that they are – *pace* the genealogical evidence – safe<sub>n</sub>.

The first of these defences is the one most frequently invoked by genealogical sceptics, especially experimental philosophers:

Our claim is neither epistemological nor metaphysical, but methodological (Swain, Alexander and Weinberg 2006, 151).

I have heard the following *tu quoque*: ‘Your arguments against appeal to intuition in philosophy are themselves grounded in intuition.’ I do not think so; I think they are grounded in psychology and successful scientific practice (Cummins 1998, 127, n. 8).

[S]ome uses of intuition, including those about logic and math, and about epistemic principles whose merits can be partially tested in the laboratory of the history of science, can reasonably be [considered reliable], and we can trust them for establishing premises to use in our arguments—including...my arguments here (Weinberg 2007, 340).

Here the sceptic’s thought seems to be: I am justified in believing and asserting the premises of my sceptical argument because they are the result of a *special method* – namely the method of good scientific practice.<sup>181,182</sup>

---

<sup>181</sup> Again, there is a striking parallel between the current debate and the aforementioned debate about the rationality of religious exclusivism. Plantinga makes a *tu quoque* charge against pluralist sceptics of religious exclusivism:

But, of course, the same goes for the pluralist. Pluralism isn’t and hasn’t been widely popular in the world at large; if the pluralist had been born in Madagascar, or medieval France, he probably wouldn’t have been a pluralist. Does it follow that he

The idea that something as various as scientific (or ‘ordinary’) practice could settle the question of whether (I6) and (I7) are true strikes me as far-fetched. Both (I6) and (I7) are fine-grained epistemic claims. Both fall out of a picture on which reliability is epistemically important, and thus evidence of unreliability at least a *prima facie* problem. That rough picture is one no doubt adopted and vindicated by good scientific practice. But it is naïve to think that that picture – a general commitment to the epistemic importance of reliability – contains the fine detail that these genealogical sceptics claim it does. Nowhere in the history of science have claims like (I6) or (I7) been tested. A broad commitment to the epistemic importance of reliability is entirely compatible with the denial of (I6), (I7) or both. To move from the generalities of ‘good scientific practice’ to the specifics of (I6) and (I7) requires epistemic judgment: just the kind of judgment the genealogical sceptic wishes to undermine.

#### 4. Rehabilitating genealogical scepticism

In the previous section I detailed how the most plausible arguments for genealogical scepticism all require various first-order epistemic premises – commitment to which is itself impugned by the respective arguments. I have also

---

shouldn’t be a pluralist or that his pluralist beliefs are produced in him by an unreliable belief-producing process? (1995, 212).

The pluralist John Hick’s response to is to draw a distinction between religious and philosophical claims, in much the same way that the experimentalist draws a distinction between epistemological and methodological claims:

[Plantinga’s] *tu quoque*, that I might well not have advocated religious pluralism if I had been born in many other times or places, and that I affirm it in much the same way that others affirm traditional Christianity, misses the all-important difference that religious pluralism...is not another religious faith or dogma alongside others, but a second-order philosophical theory, or hypothesis, about the relationship between the world religions when these are understood religiously as distinguished from naturalistically (Hick 2001, 57).

<sup>182</sup> I offer an alternative reading of these statements in section 4.1.

argued that the genealogical sceptic cannot sidestep this accusation without engaging in special pleading.

Two other strategies for rehabilitating genealogical scepticism are worth discussing. The first strategy concedes that genealogical scepticism is self-defeating, but insists that nonetheless we have pragmatic reason for accepting its conclusion. The second strategy again concedes that genealogical scepticism is self-defeating but claims that it is only intended as an *ad hominem* attack. I briefly address each strategy in turn.

#### 4.1 The pragmatic strategy

All five of the arguments I've laid out for genealogical scepticism involve a premise about the conditions under which judgments are justified. The genealogical sceptic's conclusion – that our philosophical judgments are unjustified – implies that we are not justified in believing that epistemic premise.

But the sceptic might argue that, while we would not be (epistemically) justified in believing his epistemic premise, we would nonetheless be *pragmatically* justified in acting as if it were true. How so? We can distinguish between a claim's being *true* and being *truth-conducive*. A claim is truth-conducive just in case believing it tends to result in a net gain of important true beliefs. A claim can be truth-conducive even if not true. For example, even if all mathematical claims were false (because, say, there are no numbers), employing some of them in our reasoning about the world – namely, those mathematical claims we typically take to be true – would likely result in a net gain of important true beliefs (about the natural sciences, for example). Now suppose that epistemic nihilism is true: all claims about epistemic justification are false. Nonetheless, it might be that some such claims are truth-conducive: believing them tends to result in a net gain of

important true beliefs. If we are interested in believing the true, argues the pragmatist, we should be willing to use these claims in our reasoning about the world.

Critically, genealogical scepticism about philosophical judgment impugns the truth but not the truth-conduciveness of the sceptic's epistemic premises. (Analogously, if evolution does indeed debunk morality, it might debunk the truth of our moral beliefs without debunking their conduciveness to human flourishing.) If the genealogical sceptic can offer us reason to think that his epistemic premises are truth-conducive, then it seems he can offer us pragmatic reason to act as if they are true – and thus accept his sceptical conclusion.<sup>183</sup>

What reason can the sceptic give us? He might argue that his epistemic premises have vindicated themselves as truth-conducive in the laboratories of evolution or science. This is an alternative way of reading the comments I quoted earlier by experimental philosophers:

I have heard the following *tu quoque*: 'Your arguments against appeal to intuition in philosophy are themselves grounded in intuition.' I do not think so; I think they are grounded in psychology and successful scientific practice (Cummins 1998, 127, n. 8).

[S]ome uses of intuition, including those about logic and math, and about epistemic principles whose merits can be partially tested in the laboratory of the history of science, can reasonably be [considered reliable], and we can trust them for establishing premises to use in our arguments—including...my arguments here (Weinberg 2007, 340).

On this alternative reading – one undertaken in a more revisionary spirit – the genealogical sceptic is claiming that science has proven his epistemic premises truth-conducive, independently of their truth.

---

<sup>183</sup> My suspicion is that many philosophers think along these lines, and that this explains in part why few seem very worried about the potential spread of scepticism about moral realism to scepticism about epistemic realism. On the relationship between moral realism and epistemic realism – and also criticism of 'reductionism' about epistemology (according to which epistemic norms are merely truth-conducive hypotheticals) – see Cuneo 2007.

What should we make of the pragmatic strategy for eschewing the charge of sceptical self-defeat? The problems raised in the last section equally obtain here. Evolution or science might very well vindicate the truth-conduciveness of coarse-grained epistemic principles: the value of reliability, inference to the best explanation, Occam's razor and so on. But it would be naïve to think that either evolution or science vindicated the truth-conduciveness of epistemic principles any more fine-grained than that. To say, for example, that being committed to reliability is truth-conducive does not answer the questions that are important for evaluating genealogical scepticism: *When is something evidence of unreliability? What constitutes a suitably independent method of corroboration? Is independent corroboration always necessary?* Similar things can be said of other epistemic principles that genealogical sceptics like to invoke, like the importance of respecting one's evidence or looking for good explanations. Neither evolution nor science offers an Archimedean vantage point on philosophy. Thus the pragmatic strategy for sidestepping accusations of self-defeat is not, I think, very promising.

#### 4.2 The *ad hominem* strategy

According to Sextus Empiricus, the (Pyrrhonian) sceptic does not put forward an argument in order to convince his interlocutor of its conclusion. Instead, his argument is meant to reveal an instability in the interlocutor's own accepted premises. The sceptic meanwhile simply suspends judgments on all matters, including his argument's premises and conclusion – and recommends, for the sake of your psychic wellbeing, that you do the same.

The kind of genealogical sceptics I have been discussing here do not seem to be Pyrrhonian sceptics after the model of Sextus.<sup>184</sup> But should they be? That is, should they take themselves to be merely revealing an instability in the premises accepted by their philosopher interlocutors? While I have taken pains to show that the epistemic premises plausibly required by the genealogical sceptic's argument are hardly universally accepted, it's surely the case that some philosophers accept the sceptic's premises – and are thus susceptible to an *ad hominem* attack.

What might a philosopher do once she has come to believe that her own epistemic commitments are self-undermining? Two options seem available. The first is to reject her previous epistemic commitments and take on new ones – moving to the more stable epistemic ground of anti-sceptical externalism. (This is what I mean to suggest when I speak of the 'epistemic resources' at philosophers' disposal.) But what if the philosopher cannot bring herself to adopt epistemic views congenial to dismissing genealogical scepticism? She might think, after all, that she has no reason to do so: her epistemic commitments have been revealed to be unstable, but not false. Why then believe their negation or contraries?

Perhaps such a philosopher should take up Sextus' advice and suspend judgment about the relevant epistemic premises and, with them, the genealogical sceptic's conclusion. But how, we might wonder, could such a philosopher continue to philosophise? Not only has she suspended judgment about various epistemological questions, she has moreover suspended judgment about whether any of her philosophical judgments are justified. This worry is a version of what is known as

---

<sup>184</sup> One possible exception is Robert Cummins, who writes, in response to a *tu quoque* charge against his own attack on philosophical use of 'intuition': "here is a *tu quoque* back: If you believe in intuition, and think my premises and logic are intuitive, you should accept my conclusion. If you do that, you have a *reductio* against intuition on your hands" (1998, 127, n. 8). As I noted earlier (see n. 157), Street (2006, n. 57) briefly addresses the worry that her debunking argument against evaluative realism is self-defeating. Here she seems to assume that we will in fact share her epistemic judgments – and thus, according to her metanormative constructivism, have reason to accept her conclusion. I've given some reason to think that not all philosophers do in fact share her epistemic judgments. But those that do might construe her argument charitably as an *ad hominem* attack against them.

the Apraxia Charge against Pyrrhonian scepticism. According to that charge, the sceptic who suspends judgment about everything is incapable of action, for action presupposes belief. Is it really possible for the philosopher to continue to perform philosophical actions – to entertain thought experiments, make arguments, deliberate and debate – all the while suspending judgment about all philosophical questions?

It depends of course what ‘suspending judgment’ really amounts to. According to Michael Frede (1979, 1984), the sceptics’ answer to the Apraxia Charge lies in 1.13 of *Outlines of Pyrrhonism*, in which Sextus seems to distinguish between two kinds of assent, only one of which is dogmatic and thus problematic.<sup>185</sup> Dogmatic assent, Frede claims, involves an outright belief and assertion of how things really are, while non-dogmatic assent does not. This is how we are to understand the sceptic’s claim, in acting, to being merely guided by appearances (1.21-24) and, in speaking, to be merely reporting how things seem to him (1.21-24).<sup>186</sup> Thus Frede writes that “having a view involves one kind of assent, whereas taking a position, or making a claim, involves another kind of assent, namely the kind of assent the sceptic will withhold” (1984, 128), and “[t]o be left with the impression or thought that *p*...does not involve the further thought that it is true that *p*” (ibid., 133). If Frede’s reading is correct, then perhaps the philosopher might be able to act on – indeed *philosophise* on – her culturally, evolutionary (and so on) contingent judgments, without thereby being dogmatically committed to premises that undermine that very commitment. Indeed, in his discussion of how thought is possible for the sceptic, Sextus explains that we acquire our conceptual structure not through dogmatically assenting to various propositions, but rather through cultural transmission and education; acting in accordance with this conceptual structure then does not constitute dogmatism (1.23 – 4).

---

<sup>185</sup> For a critical response, see Burnyeat 1980 and 1997.

<sup>186</sup> That is (according to Frede), the key distinction is not between beliefs with different contents (how things are versus how things appear) but rather different kinds of doxastic attitude (one of full-on dogmatic belief, and another somehow short of that).

I'm not in a position to comment on whether Frede's interpretation of Sextus is correct. But I do think its application to the question of what stance one might take on one's own philosophical judgments interesting. One cannot treat oneself as a purely psychological phenomenon; one must take up at least some of one's views from the internal perspective to be able to think and act at all. But according to Frede's Sextus, there is a belief-like attitude that is somewhere in between outright belief and total psychologisation, between a wholehearted assent to *p* and the detached observation that one is inclined to judge that *p*. If there is indeed such a space, it might provide a home for the philosopher who finds her own epistemic commitments undermining themselves.<sup>187</sup>

## 5. Resisting Archimedeanism

Explaining why he thinks traditional philosophy in peril, the experimental philosopher Joshua Alexander writes:

If our theoretical commitments shape our philosophical intuitions, it is hard to see how our philosophical intuitions can help us *independently* assess the accuracy of those theories (2012, 95).

Alexander's point is well taken: there exists no Archimedean point from which to defend philosophy as a whole. But, like other genealogical sceptics, Alexander doesn't seem to grasp the full implications of this observation; there is also no Archimedean point from which to *impugn* philosophy as a whole. Genealogical critique of philosophical judgment can and must be scrutinised from within philosophy, for there is no recourse to a philosophically neutral perspective from which to do otherwise. A genealogical critique putatively launched from outside philosophy – a 'merely methodological' critique, say, or a 'scientific' critique, or a

---

<sup>187</sup> Marcus Giaquinto suggested to me that this is roughly how we should understand Hume's metaphilosophical outlook.

critique inspired by ‘ordinary practice’ – invariably defeats itself, taking for granted various starting points that themselves constitute contested philosophical positions.

What does it mean to own up to the impossibility of an Archimedean perspective on philosophy? What sort of metaphilosophical view does such a recognition recommend? In the previous section I discussed one possible answer: a Pyrrhonian suspension of judgment about all philosophical questions. Perhaps even a more radical answer is called for. Perhaps we should dispense with the notion of a mind-independent reality altogether, as neo-pragmatists like Hilary Putnam and Richard Rorty urge. Perhaps we should think, with Rorty, that philosophising is simply about formulating creative descriptions (‘of the world’, we might incorrectly put it) that others might find useful for their practical purposes. To Rorty’s suggestion, Bernard Williams insists that philosophy must not dispense with some robust notion of ‘getting it right’:

[T]he demand to get it right has great survival value. All the philosophers who have been found interesting for more than a very brief period of fashion have been driven by a need to get it right in some terms or other. Even Nietzsche, the thinker who most self-consciously constructed himself in a new style, and most radically mistreated received standards of relevance and correctness, frequently reminded himself and any readers he might have that he was originally a philologist and had derived from that a respect for the decencies of exactness...[S]ome acknowledgment of the need is required, some concern for truthfulness that goes beyond the disposition to put next what occurs next (1989, 3).

Williams’ own view is that we can vindicate a robust notion of getting it right while taking seriously the historical situatedness of philosophy, by attending to the way in which our identities are bound up with our particular genealogies. Thus Williams writes that:

[O]nce one goes *far enough* is recognizing contingency, the problem to which [Rortyan] irony is supposed to provide the answer does not arise at all...[B]ecause we are not unencumbered intelligences

selecting in principle among all possible outlooks, we can accept that this outlook is ours just because of the history that has made it ours; or, more precisely, has made both us, and made the outlook as something that is ours...We and our outlook are not simply in the same place at the same time. If we really understand this, deeply understand it, we can be free of what is indeed another scientific illusion, that it is our job as rational agents to search for, or at least move as best we can towards, a system of political and ethical ideas which would be the best from an absolute point of view, a point of view that was free of contingent historical perspective (2000, 193-4).

For Williams, elucidating our contingent concepts and systematising our contingent judgments is no arbitrary project, for it is the project of making sense of *ourselves*.<sup>188</sup>

I don't wish to gainsay Williams' resolution of the conundrum of how to preserve a robust notion of getting it right in philosophy while acknowledging genealogical contingency. But I have here suggested that an even more robust conception of getting it right is available to us, one that is still compatible with a recognition of the genealogical contingency of philosophical judgment. The conception I have in mind is a realist one, according to which philosophy aims to describe the way the world is, independent of our conception of it. Such a realist metaphilosophy can be maintained in the light of genealogical contingency, I've suggested, by taking more seriously the role that *luck* plays in the acquisition of knowledge, including philosophical knowledge. The luck I have in mind is the luck of having the *right* genealogy.

As I argued in Essay Two, our judgments about the role that luck plays in the acquisition of knowledge are intimately bound up with our moral worldview. In particular they are bound up with the conviction, widespread in modernity, that people are only appropriately judged for those things that lie within their sphere of control. It seems to us deeply unfair to judge someone negatively for something

---

<sup>188</sup> For a broadly Williamsian metaphilosophy applied to the topic of metaphysics, see Moore 2012.

that befell her as a matter of bad luck, and similarly unfair to judge someone positively for something that happened merely as a matter of good luck. A similar instinct can drive a certain sceptical attitude toward knowledge: it would be unfair if some were blocked from acquiring knowledge because of an unfavourable genealogy, and similarly unfair if others were privy to knowledge because of some genealogical fluke. This instinct will be checked by the commonsense recognition that certain contingent things must be in place to allow for knowledge; had I never been taught math, I wouldn't know anything about calculus. But here we console ourselves with the thought that any intelligent agent, whatever her current beliefs, could be brought around to seeing how the calculus operates. We fear that the same thing cannot be said of morality, religion or philosophy; here, different genealogies produce deep, embedded and structural disagreements. If everyone is going to be equally in a position to acquire knowledge in these domains, then genealogy cannot matter so decisively. Thus the natural push to nihilism, according to which everyone is equally ignorant – or some sort of relativism, according to which everyone equally knows.

On the alternative view I've been advancing, we should abandon our firm conviction that getting it right in philosophy cannot be a matter of luck. To get it right, philosophically (as elsewhere), having the right genealogy matters, and this is never entirely within agential control. Such a metaphilosophical view can doubtless sound morally unsavoury. It is one thing to say, after all, that philosophers have a superior, 'expert' method for addressing philosophical questions – and yet another to say that Westerners (as opposed to Easterners) do. The implausibility of the latter claim isn't logical or metaphysical, but ethical: it just sounds culturally chauvinistic to say that Westerners are simply better than Easterners at knowing the philosophical truths, that they are simply equipped with a superior method. To return to an earlier example, such moral qualms play a

large role in the debate about the rationality of religious exclusivism in the face of religious diversity:

Ethically, Religious Exclusivism has the morally repugnant result of making those who have privileged knowledge, or who are intellectually astute, a religious elite, while penalizing those who happen to have no access to the putatively correct religious view, or who are incapable of advanced understanding...(Runzo 1988, 197-343).

...except at the cost of insensitivity or delinquency, it is morally not possible actually to go out into the world and say to devout, intelligent fellow human beings we believe that we know God and we are right; you believe that you know God, and you are totally wrong (Smith 1976, 14).

There is a striking resemblance here to many of the things said by experimental philosophers. For example, Weinberg, Nichols and Stich (2001, 22) ask why we should “privilege our intuitions rather than the intuitions of some other group?” and Machery et al. (2004, 9) claim that “[i]n the absence of a principled argument about why philosophers’ intuitions are superior, [analytic philosophy of language] smacks of narcissism in the extreme”.

In my view, the moralising language apparent in much genealogical scepticism is more than just rhetorical bluster. It signals a genuine discomfort with the idea that getting philosophy right could be a matter of luck – and a genuine discomfort with anyone proclaiming themselves in this way lucky. That a certain moral conviction is in play here is brought out by cases in which this conviction comes into conflict with other deeply held moral beliefs. Recall my case in which you discover your friend Robert to be an entrenched homophobe. My contention was that this case constitutes a plausible counterexample to the claim that one cannot bootstrap oneself into knowing that one’s judgment is the result of a distinct and (superior) method. Inasmuch as this case works, I think it does so because it combats one moral conviction – that access to moral knowledge isn’t a matter of genealogical

luck – with another, stronger moral conviction: that homosexuality is ethically unobjectionable.

That in some cases it seems plausible to say that getting it right *is* a matter of genealogical luck – the luck of growing up in an unbigoted household, or not being a brain-in-a-vat, or having had math lessons – is some reason to think that we should abandon the moral conviction that drives us to say that some kinds of knowledge are immune from genealogical luck. Another reason, as I discussed in Essay Two, is that we can draw a distinction between doing well (or poorly) on one hand, and being praiseworthy (or blameworthy) on the other. That some group has the good fortune of possessing a knowledge-conducive genealogy – and another the bad fortune of not – hardly makes the former praiseworthy or the latter blameworthy, anymore than I am praiseworthy for being in a position to know I have hands, and my brain-in-a-vat counterpart blameworthy for not being in such a position. Drawing this distinction goes some small way, I think, toward alleviating concerns about the moral repugnance of a view that embraces the importance of genealogical luck for philosophical knowledge.

But it does not go all the way. After all, treating someone as the unfortunate victim of poor genealogy is not the same thing as treating her with genuine respect. Nor though is simply psychologising her judgments, acting as if all there is to say is that she believes one thing, you believe another, and that is the end of it. Genuine respect demands something subtler and more complex than either of these brutish applications of realism or anti-realism. I don't wish to offer an account of what genuine respect in such a case might involve. But I do want to note that, as we try to negotiate these issues, we can and should draw another distinction: between epistemology and politics. There is on one hand the question of what it takes to know – for example, the philosophical truths – and on the other the question of how, in a just society, we should respond to other people's claims to knowledge.

This paper has been about the former question, which I take to be the far easier of the two. It is, however, a question that is perhaps too often conflated with the latter. The political conviction, essential to a just society, that everyone has the right to a fair hearing, that no ought be silenced, is often mistakenly thought to necessitate the epistemic conviction that we all have equal access to the truth. But the latter claim – since so many seek to silence others, to give no one but their familiars a fair hearing – is not consistent with the former. A just society, in other words, must take itself to have got some things right.

Nonetheless, to say that the epistemological question must be distinguished from the political one isn't to license metaphilosophical dogmatism, a brash confidence that we – whoever this 'we' is – have got it right. For philosophy too is political, a social endeavour with its own norms, hierarchy and strife, where the threat of marginalisation and silencing is all too live. (That it might be the silencing of philosophical wrongheadedness is no final defence.) In philosophy as elsewhere we are thus faced with the difficult task of cleaving to a regulative ideal of getting it right while owning up to the full political demands of our collective enterprise. Addressing that task perhaps begins with renewed attention to some prosaic virtues: empathy, openness, kindness.

## Acknowledgments

Special thanks to John Hawthorne and Tim Williamson for reading and commenting on earlier drafts of this paper. Thanks also to Troy Cross, Daniela Dover, Tanya Goodchild, Marcus Giaquinto, Andreas Mogensen, Jennifer Nagel, Daniel Rothschild and Charles Tyler for discussion of these issues. This paper is based on my BPhil thesis, “Experimental Philosophy and Armchair Philosophy” (2009) written under the supervision of John Hawthorne. I gave a talk based on the thesis at the Arché/CSMN Graduate Conference (2009); thanks to the members of that audience and especially Yuri Cath for his comments.

## Works Cited

- Aarnio-Lasonen, M. (2010). “Unreasonable Knowledge”. *Philosophical Perspectives*, 24 (1): 1-21.
- Aarnio-Lasonen, M. (forthcoming). “Higher-Order Evidence and the Limits of Defeat”. *Philosophy and Phenomenological Research*.
- Alexander, J. (2012). *Experimental Philosophy: An Introduction*. Cambridge: Polity.
- Alston, W. (1988a). “Religious Diversity and the Perceptual Knowledge of God”. *Faith and Philosophy*, 5: 433-448.
- Alston, W. (1988b). “An Internalist Externalism”. *Synthese*, 74: 265–283. Reprinted in Alston, W. (1989). *Epistemic Justification*. Ithaca, NY: Cornell University Press. [page references from reprint].
- Arendt, H. (1958). *The Human Condition*. Chicago: University of Chicago Press.
- Bealer, G. (1992). “The Incoherence of Empiricism”. *Proceedings of the Aristotelian Society Supplemental*, 66: 99-138.
- Bealer, G. (1998). “Intuition and the Autonomy of Philosophy”, in DePaul and Ramsey 1998, 201-240.
- Benacerraf, P. (1973). “Mathematical Truth”. *The Journal of Philosophy*, 70(19): 661-679.
- Bengson, J. (m.s.). “The Intellectual Given”. Available at: <https://sites.google.com/site/johnbengson/papers>
- Berlin, I. (1979). “The Concept of Scientific History”, in Hardy, H. (ed.), *Concepts and Categories*. New York: Viking Press.

Blackburn, S. (1985). "Errors and the Phenomenology of Value", In Honderich, T. and Mackie, J.L. (eds.), *Morality and Objectivity: a tribute to J.L. Mackie*. London: Routledge.

Blackburn, S. (1998). *Ruling Passions*. Oxford: Clarendon Press.

Bloomfield, P. (2009). "Archimedeanism and Why Metaethics Matters", in Shafer-Landau (ed.), *Oxford Studies in Metaethics* vol. 4. Oxford: Oxford University Press.

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

Burnyeat, M. (1980). "Can the sceptic live his scepticism?", in Shofield, M., Burnyeat, M. and Barnes, J. (eds.), *Doubt and Dogmatism: Studies in Hellenistic Epistemology*. Oxford: Oxford University Press, 20–53. Reprinted in Burnyeat and Frede (eds.) 1997, 25–57.

Burnyeat, M. and M. Frede (eds.) (1997). *The Original Sceptics*. Indianapolis: Hackett.

Chalmers, D. (2002). "Does conceivability entail possibility?" in Gendler, T. and Hawthorne, J. (eds.), *Conceivability and Possibility*. Oxford: Oxford University Press.

Chalmers, D., Manley, D. and Wasserman, R. (2009). *Metametaphysics: New Essays on the Foundations of Ontology*. Oxford: Oxford University Press.

Chisholm, R.M. (1977). *Theory of Knowledge*, 2<sup>nd</sup> edn. Englewood Cliffs, NJ: Prentice-Hall.

Christensen, D. (2007a). "Epistemology of Disagreement: the Good News". *The Philosophical Review*, 116(2): 187-217.

Christensen, D. (2007b). "Does Murphy's Law Apply in Epistemology? Self-Doubt and Rational Ideals". *Oxford Studies in Epistemology*, 2: 3-31.

Christensen, D. (2010). "Rational Reflection". *Philosophical Perspectives*, 24: 121-140.

Christensen, D. (2011). "Disagreement, Question-Begging and Epistemic Self-Criticism". *Philosopher's Imprint*, 1(6).

Coates, D. (2012). "Rational Epistemic Akrasia". *American Philosophical Quarterly*, 49(2): 113-124.

Cohen, S. (2002). "Basic Knowledge and the Problem of Easy Knowledge". *Philosophy and Phenomenological Research*, 65(2).

Conee, E. (1998). "Seeing the truth". *Philosophy and Phenomenological Research*, 58: 847-857.

Conee, E. and Feldman, R. (1998). "The Generality Problem for Reliabilism". *Philosophical Studies*, 89: 1-29.

Conee, E. and Feldman, R. (2001). "Internalism Defended". *American Philosophical Quarterly*, 38(1): 1-18.

- Cooper, W.S. (2003). *The Evolution of Reason: logic as a branch of biology*. Cambridge: Cambridge University Press.
- Copp, D. (2008). "Darwinian Skepticism About Moral Realism". *Philosophical Issues*, 18(1): 186-206.
- Cummins, R. (1998). "Reflection on Reflective Equilibrium," in DePaul and Ramsey 1998, 113 – 128.
- Cuneo, T. (2007). *The Normative Web: An Argument for Moral Realism*. Oxford: Oxford University Press.
- Darwin, C. (1879/2004). *The Descent of Man, and Selection in Relation to Sex*, 2<sup>nd</sup> ed. London: Penguin.
- Dennett, D.C. (2006). *Breaking the Spell: Religion as a Natural Phenomenon*. New York: Penguin.
- DePaul, M.R. and Ramsey, W. (eds.) (1998). *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*. Lanham: Rowman & Littlefield.
- Descartes, R. (1641/1993). *Meditations on First Philosophy*. D. Cress (trans.). Indianapolis: Hackett Publishing.
- Devitt, M. (2006). *Ignorance of Language*. Oxford: Clarendon Press.
- Devitt, M. (2011). "Experimental Semantics". *Philosophy and Phenomenological Research*, 80(2): 418-435.
- Dorr, C. (2010). "Review of Ladyman, J., and Ross, D., *Every Thing Must Go: Metaphysics Naturalized*". *Notre Dame Philosophical Reviews*.
- Dworkin, R. (1996). "Objectivity and Truth: You'd Better Believe It". *Philosophy and Public Affairs*, 25(2): 87-139.
- Elga, A. (2007). "Reflection and Disagreement". *Noûs*, 61(3): 478-502.
- Feldman, R. (2005). "Respecting the Evidence". *Philosophical Perspectives*, 19: 95-119.
- Feldman, R. (2006). "Epistemological Puzzles About Disagreement", in Hetherington, S (ed.), *Epistemology Futures*. Oxford: Oxford University Press, 216-236.
- Field, H. (1989). *Realism, mathematics and modality*. Oxford: Basil Blackwell.
- Field, H. (2005). "Recent Debates about the A Priori", in Gendler, T. and Hawthorne, J. (eds.), *Oxford Studies in Epistemology*. Oxford: Oxford University Press, 69-88.
- Feuerbach, L. (1841/1957). *The Essence of Christianity*. G. Eliot (trans.). New York: Harper & Row.
- Foley, R. (1998). "Rationality and Intellectual Self-Trust", in DePaul and Ramsey 1998.
- Frede, M. (1979). "Des Skeptikers Meinungen". *Neue Hefte für Philosophie*, 15(16): 102-29. Reprinted as "The Sceptic's Beliefs", in Burnyeat, M. and Frede, M. (eds.) 1997, 1-24.

- Frede, M. (1984). "The Sceptic's Two Kinds of Assent and the Question of the Possibility of Knowledge", in Rorty, R., Schneewind, J.B. and Skinner, Q. (eds.), *Philosophy in History*. Cambridge: Cambridge University Press, 255–78. Reprinted in Burnyeat and Frede (eds.) 1997, 127–151.
- Freud, S. (1927). *Die Zukunft einer Illusion*. Leipzig: Internationaler Psychoanalytischer Verlag.
- Gibbard, A. (1990). *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- Goldman, A. (1979). "What Is Justified Belief?", in Pappas, G., *Justification and Knowledge*. Dordrecht: Reidel.
- Goldman, A. (1986). *Epistemology and Cognition*. Cambridge, MA: Harvard University Press.
- Goldman, A. (2007). "Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status". *Grazer Philosophische Studien*, 7.
- Goldman, A. (2010). "Philosophical Naturalism and Intuitional Methodology: the Romanell Lecture". *Proceedings and Addresses of the American Philosophical Association*, 84(2): 115–150.
- Gopnik, A. and Schwitzgebel, E. (1998). "Whose concepts are they anyway? The role of philosophical intuition in empirical psychology", in DePaul and Ramsey 1998.
- Greco, D. (forthcoming). "A Puzzle About Epistemic Akrasia". *Philosophical Studies*.
- Greene, J. (2008). "The Secret Joke of Kant's Soul", in Sinnott-Armstrong, W. (ed.), *Moral Psychology, Vol. 3: The Neuroscience of Morality, Emotions, Brain Disorders and Development*. Cambridge, MA: MIT Press, 35–80.
- Gutting, G. (1982). *Religious Belief and Religious Skepticism*. Notre Dame: University of Notre Dame Press.
- Harman, G. (1977). *The Nature of Morality: An Introduction to Ethics*. New York: Oxford University Press.
- Harman, G. (1986). "Moral explanations of natural facts: can moral claims be tested against moral reality?" *Southern Journal of Philosophy*, 24: 57–68.
- Hawthorne, J. and Srinivasan, A. (2013). "Disagreement without Transparency: some bleak thoughts". In Christensen, D. and Lackey, J. (eds.), *The Epistemology of Disagreement*. Oxford: Oxford University Press, ch. 1.
- Hegel, G.W.F. (1820/1991). *Elements of the Philosophy of Right*. Wood, A.W. (ed.) and Nisbet, H.B. (transl.) Cambridge: Cambridge University Press.
- Hegel, G.W.F. (1837/2001). *The Philosophy of History*. Sibree, J. (ed. and transl.) with a preface by Charles Hegel. Batoche Books: Kitchener, Ontario.
- Hick, J. (1980). "Towards a Philosophy of Religious Pluralism". *Neue Zeitschrift für Systematische Theologie und Religionsphilosophie*, 22.

- Hick, J. (1982). *God Has Many Names*. Louisville, PA: Westminster.
- Hick, J. (1989). *An Interpretation of Religion: Human Responses to the Transcendent*. New Haven: Yale University Press.
- Hick, J. (2001). *Dialogues in the Philosophy of Religion*. New York: Palgrave.
- Horowitz, S. (forthcoming). “Epistemic Akrasia”. *Noûs*.
- Huemer, M. (2007). “Compassionate phenomenal conservatism”. *Philosophy and Phenomenological Research*, 74: 30-55.
- Huemer, M. (2008). “Revisionary Intuitionism”. *Social Philosophy and Policy*, 25: 368-392.
- Hurley, S. (1989). *Natural Reasons: personality and polity*. Oxford: Oxford University Press.
- Joyce, R. (2006). *The Evolution of Morality*. Cambridge, MA: MIT Press.
- Kauppinen, A. (2007). “The rise and fall of experimental philosophy”. *Philosophical Explorations*, 10: 95-118.
- Kitcher, P. (2005) “Biology and Ethics,” in Copp, D. (ed.), *The Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.
- Kitcher, P. (2011). *The Ethical Project*. Cambridge: Harvard University Press.
- Knobe, J. and Nichols, S. (2008). *Experimental Philosophy*. Oxford: Oxford University Press.
- Knobe, J. and Nichols, S. (forthcoming). *Experimental Philosophy, Vol. 2*. Oxford: Oxford University Press.
- Kolodny, N. (2005). “Why Be Rational?” *Mind*, 114(455): 509 - 563
- Kramer, M. (2009). *Moral Realism as a Moral Doctrine*. Oxford: Blackwell.
- Ladyman, J. and Ross, D., with Spurrett, D. and Collier, J. (2007). *Every Thing Must Go: Metaphysics Naturalized*. Oxford: Oxford University Press.
- Lewis, D. (1973). *Counterfactuals*. Oxford: Blackwell.
- Liao, M. (2008), “A defense of intuitions”. *Philosophical Studies*, 140: 247-262.
- Lipton, P. (2004). “What Good is an Explanation”, in Cornwell, J. (ed.), *Explanations: Styles of Explanation in Science*. Oxford: Oxford University Press, 1-21.
- Machery, E., Mallon, R., Nichols, S. and Stich, S. (2004). “Semantics: Cross-Cultural Style”. *Cognition*, 92: B1-B12. Reprinted in Knobe and Nichols 2008.
- Marx, K. (1844/1975). “A Contribution to the Critique of Hegel's Philosophy of Right”, in Livingstone, R.A.B. (ed.), *Marx: Early Writings*. London: Penguin Books.

- McGinn, C. (2012). *Truth by Analysis: Games, Names, and Philosophy*. Oxford: Oxford University Press.
- Merricks, T. (2001). *Objects and persons*. Oxford: Clarendon Press.
- Mogensen, A. (m.s.). "Evolution, Ethics, and the Coincidence Problem: A Skeptical Appraisal".
- Moore, A.W. (2012). *The Evolution of Metaphysics*. Cambridge: Cambridge University Press.
- Nagel, J. (2012). "Intuitions and experiments". *Philosophy and Phenomenological Research*, 85(3): 495-527.
- Nagel, T. (1996). *The Last Word*. Oxford: Oxford University Press.
- Nagel, T. (2012). *Mind and Cosmos*. Oxford: Oxford University Press.
- Nerlich, G. (1979). "What Can Geometry Explain?" *British Journal for the Philosophy of Science*, 30: 69-83.
- Nietzsche, F. (1881/1911). *The Dawn of Day*. Kennedy, J.M. (transl.). New York: MacMillan.
- Nietzsche, F. (1887). *Zur Genealogie der Moral*. Leipzig: C.G. Naumann.
- Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Pinillos, A.N., Smith, N., Nair, S., Marchetto, P. and Mun, C. (2011). "Philosophy's New Challenge: Experiments and Intentional Action". *Mind & Language*, 26(1): 115-139.
- Plantinga, A. (1986). "The Foundations of Theism: A Reply." *Faith and Philosophy*, 3(3): 310-312.
- Plantinga, A. (1993). *Warrant and Proper Function*. New York: Oxford University Press.
- Plantinga, A. (1994). "Naturalism Defeated". Published online at: [http://www.calvin.edu/academic/philosophy/virtual\\_library/articles/plantinga\\_alvin/naturalism\\_defeated.pdf](http://www.calvin.edu/academic/philosophy/virtual_library/articles/plantinga_alvin/naturalism_defeated.pdf)
- Plantinga, A. (1995). "A Defense of Religious Exclusivism", in Senor, T.D. (ed.), *The Rationality of Belief and The Plurality of Faith*. Ithaca: Cornell University Press.
- Plantinga, A. (2000). "Pluralism: A Defense of Religious Exclusivism", in Meeker, K. and Quinn, P. (eds.), *The Philosophical Challenge of Religious Diversity*. New York: Oxford University Press, 172-192.
- Plantinga, A. (2002). "The Evolutionary Argument Against Naturalism", in Beilby, J. (ed.), *Naturalism Defeated?* Ithaca: Cornell University Press, 1 - 12.
- Prinz, J. (2008). "Empirical Philosophy and Experimental Philosophy", in Knobe and Nichols 2008, 189-208.
- Pust, J. (2000). *Intuitions as Evidence*. New York: Garland Publishing.

- Pust, J. (2001). "Against explanationist skepticism regarding philosophical intuitions". *Philosophical Studies*, 106: 227-58.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Rawls, J. (1974). "The Independence of Moral Theory". *Proceedings and Addresses of the American Philosophical Association*, 48.
- Rosenberg, A. (2011). *The Atheist's Guide to Reality: Enjoying Life Without Illusions*. New York: W. W. Norton.
- Runzo, J. (1988). "God, Commitment, and Other Faiths: Pluralism vs. Relativism". *Faith and Philosophy*, 5: 343-364.
- Ruse, M. (1985). *Taking Darwin Seriously*. Oxford: Blackwell.
- Ruse, M. and Wilson, E.O. (1986). "Moral Philosophy as Applied Science". *Philosophy*, 61: 173-92.
- Sextus Empiricus (1933). *Outlines of Pyrrhonism*. Loeb Classical Library. Cambridge, MA: Harvard University press.
- Sider, T. (forthcoming). "Against Parthood", in Bennett, K. and Zimmerman, D. (eds.), *Oxford Studies in Metaphysics, volume 8*. Oxford: Oxford University Press.
- Silver, D. (2001). "Religious Experience and the Facts of Pluralism". *International Journal for the Philosophy of Religion*, 49: 1-17.
- Singer, P. (1981). *The Expanding Circle*. New York: Farrar, Straus & Giroux.
- Singer, P. (2005). "Ethics and Intuitions". *Journal of Ethics*, 9: 331-52.
- Smith, W.C. (1976). *Religious Diversity*. New York: Harper and Row.
- Smithies, D. (2012). "Moore's Paradox and the Accessibility of Justification". *Philosophy and Phenomenological Research*, 85(2): 273-300.
- Sober, E., (1983). "Equilibrium Explanation". *Philosophical Studies*, 43: 201-210.
- Sosa, E. (1996). "Rational intuition: Bealer on its nature and epistemic status". *Philosophical Studies*, 81:151-162.
- Sosa, E. (1998). "Minimal intuition", in DePaul and Ramsey 1998.
- Sosa, E. (1999). "How to defeat opposite to Moore". *Noûs*, 33: 141-153.
- Sosa, E. (2005), "A Defense of the Use of Intuitions in Philosophy", in Bishop, M. and Murphy, D. (eds.), *Stich and His Critics*. Oxford: Blackwell.
- Sosa, E. (2007). "Experimental Philosophy and Philosophical Intuition". *Philosophical Studies*, 132(1): 99-107.
- Street, S. (2006). "A Darwinian Dilemma for Realist Theories of Value". *Philosophical Studies*, 127: 109-166.

- Street, S. (2008a). "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About". *Philosophical Issues*, 18: 207-228.
- Street, S. (2008b). "Constructivism about Reasons", in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics, Volume 3*. Oxford: Oxford University Press, 207-246.
- Street, S. (2011). "Mind-Independence Without Mystery: Why Quasi-Realists Can't Have It Both Ways", in Shafer-Landau, R. (ed.), *Oxford Studies in Metaethics, Volume 6*. Oxford: Oxford University Press, 1-32.
- Sturgeon, N. (1988). "Moral Explanations", in Sayre-McCord, G. (ed.), *Essays on Moral Realism*. Ithaca: Cornell University Press, 229-255.
- Swain, S., Alexander, J. and Weinberg, J. (2006). "The Instability of Philosophical Intuitions: Running Hot and Cold on Truetemp". *Philosophy and Phenomenological Research*, 76(1).
- Tidman, P. (1996). "The Justification of A Priori Intuitions". *Philosophy and Phenomenological Research*, 56(1): 161 – 171.
- Titelbaum, M. (forthcoming). "Rationality's Fixed Point". *Oxford Studies in Epistemology*.
- van Cleve, J. (2003). "Is Knowledge Easy – or Impossible? Externalism as the Only Alternative to Skepticism," in Luper, S. (ed.), *The Sceptics*. Aldershot: Ashgate.
- Vogel, J. (2008). "Epistemic Bootstrapping". *Journal of Philosophy*, 105(9): 518-539.
- Weatherson, B. (m.s.). "Do Judgments Screen Evidence?" Available at: <http://brian.weatherson.org/JSE.pdf>
- Wedgwood, R. (2002). "Internalism Explained". *Philosophy and Phenomenological Research*, 65(2): 349-369.
- Wedgwood, R. (2006). "The Normative Force of Reasoning". *Noûs*, 40(4): 660-686.
- Wedgwood, R. (2007). *The Nature of Normativity*. Oxford: Oxford University Press.
- Wedgwood, R. (2011). "Justified Inference". *Synthese*, 189(2): 1-23.
- Weinberg, J. (2007). "How to Challenge Intuitions Empirically Without Risking Skepticism". *Midwest Studies in Philosophy*, 31(1):318-343.
- Weinberg, J., Nichols, S. and Stich, S. (2001). "Normativity and Epistemic Intuitions". *Philosophical Topics*, 29: 429-460. Reprinted in Knobe and Nichols 2008, 17-45.
- Weinberg, J., Nichols, S. and Stich, S. (2003). "Metaskepticism: Meditations in Ethno-Epistemology", in Luper, S. (ed.), *The Sceptics*. Aldershot: Ashgate, 227-247.
- White, R. (2006.) "Problems for Dogmatism". *Philosophical Studies*, 131(3): 525-57.
- White, R. (2010). "You just believe that because..." *Philosophical Perspectives*, 24(1): 573-615.

- Willard, J., (2001). "Alston's Epistemology of Religious Belief and the Problem of Religious Diversity." *Religious Studies*, 37: 59-74.
- Williams, B. (1985). *Ethics and the Limits of Philosophy*. Cambridge, MA: Harvard University Press.
- Williams, B. (1989). "Getting it Right". *London Review of Books*, 11(22): 3-5.
- Williams, B. (2000). "Philosophy as a Humanistic Discipline". *Philosophy*, 75. Reprinted in Moore, A.W. (ed.), (2006), *Philosophy as a Humanistic Discipline*. Princeton: Princeton University Press. 180-199. [page references from reprint].
- Williams, B. (2002). *Truth and Truthfulness*. Princeton: Princeton University Press.
- Williamson, T. (2000). *Knowledge and Its Limits*. Oxford: Oxford University Press.
- Williamson, T. (2004a). "Philosophical 'Intuitions' and Skepticism about Judgement". *Dialectica*, 58(1): 109 – 53.
- Williamson, T. (2004b). "The Presidential Address I—Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking". *Proceedings of the Aristotelian Society*, 105(1): 1-23.
- Williamson, T. (2007). *The Philosophy of Philosophy*. Oxford: Blackwell.
- Williamson, T. (2011). "Improbable Knowing", in Dougherty, T. (ed.), *Evidentialism and its Discontents*. Oxford: Oxford University Press, ch. 9.
- Williamson, T. (forthcoming). "Review of J. Alexander, *Experimental Philosophy: An Introduction*". *Philosophy*.
- Woodward, J., (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Xenophanes (1898). "Fragments", in Fairbanks, A. (ed. and trans.), *The First Philosophers of Greece*. London: K. Paul, Trench, Trubner.
- Yablo, S. (1993). "Is conceivability a guide to possibility?" *Philosophy and Phenomenological Research*, 53: 1-42.