

**THE METHOD OF MORAL HYPOTHESIS**

by

**Stephen C. Makin**

B.A. in Philosophy, University of Chicago, 2003

Submitted to the Graduate Faculty of  
the Kenneth P. Dietrich School of  
Arts and Sciences in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2012

UNIVERSITY OF PITTSBURGH  
DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Stephen C. Makin

It was defended on

August 23, 2012

and approved by

Kieran Setiya, Associate Professor of Philosophy

Peter Machamer, Professor of the History and Philosophy of Science

John McDowell, Distinguished University Professor of Philosophy

Dissertation Advisor: Robert Brandom, Distinguished Professor of Philosophy

Dissertation Advisor: Michael Thompson, Professor of Philosophy

Copyright © by Stephen C. Makin

2012

## THE METHOD OF MORAL HYPOTHESIS

Stephen C. Makin, PhD

University of Pittsburgh, 2012

Moral philosophy has become interested again in particular, substantive questions of right and wrong. In an effort to divine answers to such questions, philosophers often employ the following method: general rules are floated as potential principles of morality; the principles are regarded as confirmed insofar as they match our pre-theoretical intuitions about particular cases; and otherwise infirmed. Such principles, if *sufficiently* confirmed, are then used to overturn other, ‘aberrant’ moral intuitions that do not square with the rule.

The aim of this work is to indict this ‘method of moral hypothesis’, and with it the *moral theory* project which relies on it. I argue that the method trades on an unsustainable picture of moral epistemology; that the motivations for engaging in it are without merit; and that its attractions as a systematizing tool are illusory.

In chapter one, I examine some recent ‘etiological’ skeptical challenges to moral knowledge; and argue that such challenges succeed only against a particular sort of moral epistemology—the kind to which the moral theory project is wedded. I conclude that we should reject this epistemology, and the project with it.

Chapter two aims to vindicate the charges of *Pessimists* about moral testimony—those who claim that testimony cannot transmit moral knowledge. I argue that one barrier to moral-knowledge transmission by testimony is its inability to transfer moral-conceptual ‘know-how’; more generally that the ‘Humean reasons’ which support testimony are insufficient to support

moral knowledge; and that, for parallel reasons, the theory project cannot produce moral knowledge.

Chapter three attacks a picture of justification which makes the theory project seem pressing. In its place, I argue for an alternative picture, on which justification is infected with certain *pragmatic, contextual* factors. This alternative undermines one of the motivations for the theory project: finding an *ultimate justification* for our moral beliefs.

In chapter four, I unify these arguments; and argue that, in general, we are correct to reject any summarizing principle which conflicts with a strongly held, pre-theoretical moral verdict. This negates one of the central ambitions of the theory project. Its other motivations are, I argue, equally misplaced.

## TABLE OF CONTENTS

<b>PREFACE.....</b>	<b>IX</b>
<b>INTRODUCTION.....</b>	<b>1</b>
<b>1 EVIDENCE AND ETHICAL KNOWLEDGE.....</b>	<b>7</b>
1.....	7
2.....	9
3.....	14
4.....	15
5.....	17
6.....	20
7.....	24
8.....	27
9.....	29
<b>2 MORAL TESTIMONY? .....</b>	<b>32</b>
1.....	32
2.....	34
3.....	35
4.....	39
5.....	42

6.....	44
7.....	47
8.....	48
9.....	52
10.....	54
<b>3 JUSTIFICATION IN CONTEXT.....</b>	<b>56</b>
1.....	56
2.....	60
3.....	63
4.....	65
5.....	67
6.....	70
7.....	73
8.....	74
9.....	78
<b>4 THE METHOD OF MORAL HYPOTHESIS.....</b>	<b>80</b>
1.....	80
2.....	84
3.....	86
4.....	90
5.....	94
6.....	98
7.....	101

8.....	104
9.....	105
10.....	107
<b>BIBLIOGRAPHY .....</b>	<b>110</b>

## PREFACE

This work would not have been possible without the help of my Advisors, Robert Brandom and Michael Thompson—whose work has inspired me; and the contributions of my other Committee members: John McDowell, Kieran Setiya, and Peter Machamer—who kept me on the straight and narrow.

I should like also to thank my teacher, Karl Schafer, with whom I spent many philosophical hours; Eli Witus, who got the whole thing going; and my classmates, whom I was very lucky to have—most especially: my friend, Tim Willenken; but certainly also: Casey Doyle, Jack Fortune, Joe Millburn, and Eric Palmer.

Everything that is worthless in this work is my doing; everything of value, I owe to JGM—without whom not.

## INTRODUCTION

My task in the following four essays is to try cast some suspicion on the project that has sometimes been called ‘moral theory’: the project of trying to bring our various spontaneous moral judgments into some systematic, unified whole, typically, by bringing individual judgments under principles, and then those principles under more general principles still; and in particular the version of this project in which it is accepted that some of our spontaneous moral judgments will have to go, under pressure from the systematizing requirement. We find this project going on, from *one* end, in theorists who begin by deploying some very general formula, a kind of master principle of morality, and try to work their way down to particular judgments; these theorists are generally acknowledged to be engaged in ‘moral theory’. But the term—‘moral theorizing’—could be applied with equal justice to those theorists who work from the other end: who begin with some set of spontaneous moral verdicts (‘moral intuitions’) about ostensibly similar cases, and try to develop some principle which will divide the cases according to our intuitions. What is again vital, for my point, is the willingness of these theorists to abandon some recalcitrant ‘intuition’ if it fails to fit the general pattern. Of course, the offending intuition is not merely abandoned; some attempt is always made to ‘account’ for it. But, as I shall argue, there remains something problematic here nonetheless. My aim, in what follows, will be to suggest some grounds for Annette Baier’s suspicion of “the whole idea of a moral ‘theory’ which systematizes and extends a body of moral judgments, and [...] in particular the

idea that a theorist might accept a theory with controversial implications, without thereby becoming a moral reformer, one dedicated to having these implications endorsed and acted upon.”

\* \* \*

Although, in the course of these essays, I give various arguments against the theory project, the thread which connects the various arguments is the claim that that project relies on a doubtful, quasi-empirical epistemology. We can begin to get a picture of that epistemology in the following way.

When philosophers want to assert something about objectivity, we sometimes find them falling back on such phrases as ‘out there’ (“the facts are out there, independent of us”) or ‘there anyway’ (“responding to what is there anyway”). The demonstrative is in one way unfortunate, since it seems to place these facts somewhere in *space*. But although in a way also misleading, this manner of speaking actually suggests itself in the empirical case: the facts about the stars are ‘out there’ in the sense that *the stars* are out there, somewhere in space. (The tendency to conflate the *facts* and the *objects that those facts are about*, is not itself entirely innocent.) This manner of speaking is conducive, moreover, to a certain picture of how we come to *know* the facts—roughly, we are in causal contact with them. (But see the parenthetical remark; philosophers sometimes write this way, though it seems clear enough that *light* does not *bounce off* the facts.)

Whatever the merits or demerits of this way of speaking in the empirical realm, it can only engender confusion, I think, when speaking of ethics. Of course we want to insist that we don’t get to “make up the facts”; that the facts about what we ought to do, do not depend entirely on (e.g.) what we happen to want (though of course some philosophers have, with sedulous

consistency, drawn this conclusion)—in a word, that the facts are ‘objective’ or ‘real’. On the other hand, the language suggests again a *causal* relation to the facts; and here, in ethics, the idea is surely absurd. It might now look as though the ethical facts—that is, the facts about what we ought to do—could have remained—could remain even now—forever hidden from human beings, by chance or by their nature beyond our ken.

That there is an essential disanalogy—in point of “out there-ness,” as we might put it—between the facts of empirical nature and the facts of ethics is shown, I think, by very different ways in which we justify these facts, and query claims about them. The point can be summed up by saying that, if I want to query some empirical claim of yours—“She is in the library”—I shall ask, “How do you know?”; whereas if I want to query a moral claim—“Abortion is wrong”—I shall ask “*Why?*”, and never “How do you know?” Likewise, to *justify* some empirical claim, I shall, typically, advert either to a) the mode of apprehension—“I saw her,” “I heard her”—or else b) *evidence*; whereas to justify an ethical claim I shall give *reasons*.

What is the difference between ‘evidence’ and ‘reasons’, as I am using those terms? One way to put it might be this: if x, y, and z are *reasons*, then they are *reasons why P is true*. And if I come to see that P is true by being given the reasons for P, then I see that P is true *in virtue of* the reasons. Nothing like this could be said of evidence. If x, y, and z are *evidence* for P, then they are not reasons why P is true but grounds *for believing* P. And if I come to believe P on the basis of x, y, and z, I still do not see that P is true *in virtue of* x, y, and z.

This may seem a very fine distinction indeed, but it is robust, I think, when we look at cases. If I say that there is a raccoon in the house, and you query that claim, I might advert to the little (disgusting) paw-prints and the overturned garbage can. I believe, then, that there is a raccoon in the house—as you might come to believe it—on the basis of various bit of evidence;

but of course *that there are tiny paw-prints* couldn't be said to be the 'reason why' there is a raccoon in the house. Whereas, on the other hand, if I say that it would be wrong to tell Jones that story, and I justify that claim by noting that it would hurt her feelings, then *that it would hurt her feelings* is precisely the reason why it would be wrong to tell her.

The point is not limited to ethics. In many places we give *the reasons why* something is so, as opposed to 'reasons to believe': we might say why a joke is funny; why a book is a good book; or why we should go to the party tonight. Any claim made in these areas—humor, aesthetics, prudence—is queried by asking "Why?" and never "How do you know?" For this reason, it would be equally absurd to ask for *evidence* that a joke was funny, or that a book was good, and so on.

All this is connected with the language of "out there-ness" in the following way. When we justify an empirical claim, we make reference, in one way or another, to the *source* of our knowledge. For in referencing a mode of apprehension (sight, hearing), we give our interlocutor to understand how there might be a *path*, as we could put it, between ourselves and what our belief is about. In short, we make room for a *causal story*—in the broadest sense of that term—about how we could have acquired our knowledge. And when we give evidence we do something similar. When we cite evidence—paw-prints, strewn garbage—we offer our claim ("There is a raccoon in here") as the *best explanation* for these various items. That is: we justify our major claim by making room for a causal story, that runs from (say) the raccoon, through paw-prints, to ourselves. So there is a recognizable sense, in either case, in which the facts are 'out there', and then discovered by us, either by direct apprehension, or as mediated by various bits of evidence.

Whereas the ethical facts—if there are any—are not anywhere in space; and so to justify claims about them will never make reference to their source—not to a mode of apprehension nor to evidence for those claims. What we do give are *reasons*—that is, reasons why they are so. —Just as we give reasons why a joke is funny, reasons why a book is good, and reasons that we should go to the party.

\* \* \*

What I argue, in the following pages, is that the ‘main method’ of ethics, as I have described it, depends crucially on treating various moral beliefs—and specifically what are often called our ‘moral intuitions’—as bits of evidence—in effect, a quasi-empirical model of ethics.

In Chapter 1 I argue that this model for moral epistemology opens the door for fatal skeptical problems, and for that reason, should be abandoned. By treating the relationship between ethical grounds and conclusions as *evidential*, the (manifest) facts about the contingent origins of our beliefs will rush in to undermine our claims to moral knowledge. In effect, we hold our ethical beliefs hostage to an inference to the best explanation—an inference which will always be fatal, since the idea of causal contact with the ethical facts is a non-starter.

Chapter 2 approaches the problem from a different angle; it comprises some reflections on *moral testimony*. In this chapter my aim is to substantiate the charge, argued by many theorists, that there is something specially problematic about the idea of *moral* knowledge being transmitted by testimony. My thought is that one reason moral knowledge cannot be so transmitted is that such knowledge is, in part, a kind of ‘know-how’—conceptual know-how—which, by its nature, will not be conveyed by anything properly called ‘testimony’—not even if that testimony includes also the *grounds* for the major claim. But the larger point is that the

various barriers to knowledge-transmission by testimony have parallels in the methods of moral theorizing: it is equally unequipped to deliver moral knowledge.

Chapter 3 is a departure. One reason that philosophers have felt the *need* for a moral theory has been their sense that moral claims must be *justified* by further—often: broader, more general—moral claims; and these in turn by further, more general principles still. The result is the pyramidal structure of moral theory, familiar from Kant and Mill, but also more contemporary theorists—Scanlon or Parfit. My point in this chapter is to undermine the picture of justification which supports this felt need; and thereby to remove one of the major incentives for ambitious moral theorizing.

In the final chapter, I tie these themes together, and apply their lessons to the ‘main method’. I conclude that there is something very doubtful about the project’s central ambition: revising and extending our considered moral judgments about particular, local cases; and I give some grounds for skepticism for other motivations for the project.

## 1 EVIDENCE AND ETHICAL KNOWLEDGE

1. “And it never troubles him that mere accident has decided which of these numerous worlds is the object of his reliance, and that the same causes which make him a churchman in London would have made him a Buddhist or a Confucian in Peking.”<sup>1</sup> So wrote John Stuart Mill, with evident dissatisfaction, of our complacency in the face of the contingency of our most deeply held beliefs. And yet, in our more reflective moments, thoughts of this kind can and *do* strike a disquieting note. But why, exactly? Well, one wants to say, if my beliefs were in this way *contingent*; if I might’ve just as easily believed something else—then how can I have any confidence in what I believe at all?

We may indeed want to say this; and I think there is an undeniably disturbing *psychological* force to such observations; but it appears glib answers won’t do here. A little reflection reminds us that we have *all sorts* of beliefs which are, in this sense, contingent, and being reminded of their contingency sets off no alarms at all: it is a ‘contingent’ fact, I suppose, that I believe that the earth is round—others have denied it; had I been raised in different circumstances, I would have too. But for all that I feel pretty complacent: I saw the pictures.

The worry, then, tends to center on beliefs that cannot be so easily given ‘etiological credentials’: my normative beliefs; more narrowly, my moral convictions. We can hardly advert, here, to some uncontroversial mode of direct epistemic access; evidently, I learned my morals

---

<sup>1</sup> J.S. Mill, *On Liberty*, p. 17.

from my parents; or perhaps those beliefs are—in some sense—the ‘product’ of my evolutionary heritage. Neither etiology seems particularly comforting; neither seems to promise ‘direct access’ to the moral facts. Why, then, should I be so confident in these beliefs—especially when there are a hundred other systems of belief that (it appears) I might as easily have had?

Worries of this kind have recently come into focus in analytic moral philosophy.<sup>2</sup> And though some have argued that the worry affects a wider range of views, the most obvious target of ‘etiological’ challenges is moral realism.<sup>3</sup> In particular, the problem seems worst for forms of *non-naturalist moral realism*: those views which hold that the ethical facts are ‘mind-independent’, but deny that such facts are ‘causally efficacious’.<sup>4</sup> In what follows, I hope to do the following: outline a clear version of the challenge; explain which forms of moral realism it does indeed cause serious problems for; and then finally to vindicate a picture of ethics which has every right to the title ‘non-naturalist moral realism’ and yet nevertheless escapes the puzzle. But if the arguments below are right, then there are other serious consequences as well: the assumptions which make certain forms of moral realism vulnerable to the challenge turn out to be very widely held, and perhaps central to other philosophical ambitions.

---

<sup>2</sup> See: Sher 2001; Street 2006, Street n.d.; Enoch 2010; White 2010; Schafer 2010; Joyce 2006; Singer 2005. In fact, these philosophers may be worried about slightly different problems; that itself is a matter of some controversy (see White 2010). Enoch maintains that “no realist...has ever addressed the challenge” as he understands it; White notes (fn. 4) that the topic has received “surprisingly little sustained attention from analytic epistemologists.” So while the problem is old, interest in it, in the analytic tradition, appears to be new. White suggests that the problem has received more attention in ‘continental’ philosophy; see Leiter 2004, for a discussion. See also Williams 2002, pp. 219-24, for a discussion of a related problem: the threat supposedly posed by the historical origins of inequitable systems of justice.

<sup>3</sup> Enoch speaks of the broader category of ‘metanormative realism’, but it is clear from his discussion that moral realism is part of the target. Street, for her part, thinks that the worry affects quasi-realists, too (see Street, n.d, p.7); I am less sure of this, but I leave the point aside in this essay.

<sup>4</sup> Enoch calls such a view ‘robust’ realism (pp. 414-15). I take my term over from Street, who uses ‘non-naturalist (normative) realism’ in a similar way (see n.d., p. 6). Nevertheless, some self-styled non-naturalists (e.g., Oddie 2005) defend forms of non-reductive moral realism which countenance causally efficacious moral facts. Obviously, the *label* (‘non-naturalist moral realism’) is not important; I use it, in any case, only to block in, for the moment, certain familiar, appealing views, on which moral facts are indeed causally inefficacious. But some of the details of just what counts as ‘realism’ *do* matter, and I offer a more precise analysis in §§4-5 below.

2. Now the puzzle does indeed have a familiar shape, roughly the one outlined just above. But getting clear on exactly what it is—and what it is not—may be a little tricky. Sharon Street has recently forwarded a forceful version of the puzzle under the title of a ‘Darwinian Dilemma’ for moral realists.<sup>5</sup> Her challenge, in effect, is this: the best explanation for our having the moral beliefs we do is some story about evolution, for, at least at the most basic level, evolution has been an “enormous factor in shaping the content of human values.”<sup>6</sup> But (the argument continues) there is no reason to suppose that evolutionary pressures have been ‘tracking’ the normative facts—worse, it’s very difficult to see *how* they could do any such tracking if such facts are held to be ‘causally inert’. It appears, then, that our best explanation of how we arrived at our moral beliefs is one which gives no aid or comfort to our confidence in those beliefs; if the body of our moral beliefs *is* largely correct, that could only be by enormous coincidence. Street’s conclusion is that we should give up on Realism altogether, on pain of embracing skepticism, and plump for some form of anti-realism.<sup>7</sup>

But *is* evolution really the ‘best explanation’ for our moral beliefs? There is something rather doubtful about the idea. What seems at least plausible is this: had we enjoyed a different evolutionary history, we would have different moral beliefs. But surely something parallel can be said about all sorts of elements ‘upstream’ of our contemporary moral beliefs: had such-and-such religious figure not appeared on Earth, we (or anyway, a lot of us) would have very

---

<sup>5</sup> Street aims her arguments, in her 2006, at our ‘evaluative beliefs’ and in her n.d. at our ‘normative beliefs’. I have no quarrel with these terms, but my interest is in *moral* epistemology; and whatever subtleties may divide the various terms, it is quite clear that her challenge is meant to apply to realists about the ‘moral’ realm as well.

<sup>6</sup> Street 2006, 114. In fact, Street writes ‘natural selection’, rather than evolution, which, as White points out (p. 586) may be rather dubious. As I note below in the text, however, controversies over such subtleties would amount to a tempest in a teapot.

<sup>7</sup> Street 2006, especially 126-8. Street’s favored form of anti-realism is something she calls ‘constructivism’, to which I return below. In fact, Street contends that the *naturalist* realist is no better off. Her thought is that the naturalist realist is in any case committed to the existence of certain natural-normative identities; and that the argument mounted against the non-naturalist’s moral facts applies with equal force to such identities.

different moral beliefs; more humbly, had *I* had a different upbringing, or different peers, I would now have rather different moral beliefs.

Part of the force, I think, of evolutionary debunking arguments, is their insinuation that some element in our history *controls* our beliefs.<sup>8</sup> But the mere truth of some counterfactual claim, to the effect that, had things been different, I would not believe what I now do, could not by itself undermine the possibility of *rational control* over that belief. (Or else we are dealing with a much more general form of skepticism.<sup>9</sup>) But even if we waive this difficulty, it is not *prima facie* obvious that evolution has even been more ‘influential’ than various other causal factors. What we would need to establish the stronger claim is, I think, a further premise to the effect that our moral beliefs are largely insensitive to other perturbations in our causal histories: a different upbringing (e.g.) would *not* result in radically different moral views.<sup>10</sup>

Might that be true? I doubt it myself; an argument to that effect would in any case be rather speculative. The etiological skeptic is fond of reminding us of the huge variety of possible moral belief systems that people might hold; he may be less quick to note that, in imagining such people, we are presumably imagining other *products of evolution*. If such thought experiments are to be trusted, it appears that our evolutionary history is compatible with a very large variety of ethical outlooks; so the idea of evolution ‘influencing’ our moral beliefs must be

---

<sup>8</sup> Street speaks repeatedly of evolution “shaping” our beliefs in her 2006 (pp. 109, 110, 113, 117, 121, *et alibi*) and of what evolution “led us” to believe in her n.d (pp. 12, 14, 17, *et alibi*). She draws attention to some difficulties with the latter term at p. 12, fn. 18, of the latter.

<sup>9</sup> If we accepted such a principle, then the truth of determinism would be enough to show that we have rational control over *none* of our beliefs. For, surely, there is some element, in the causal history of every belief, a) for which some such counterfactual is true, and b) which is not itself a *reason to believe*. It would follow that *every* belief is ‘controlled’ by some rationally irrelevant causal factor. An illuminating discussion of this and nearby issues can be found in Lavin 2004, especially pp. 443-446. See also White, p. 574.

<sup>10</sup> Oddly, Street seems to think that she needs only the weaker, counterfactual-dependence condition: see her 2006 p.120 and fn. 21.

correspondingly weakened. (Or when we imagine the ‘coherent Caligula’, are we somehow imagining someone who is *not* the product of evolutionary forces?)

Perhaps it will be felt that our evolutionary history at least makes certain moral views *more likely* to be held by human beings. Something like that does seem plausible. (On the other hand, we might consider the variety of moral views held in different eras and different places.)<sup>11</sup> But it is hard to know exactly how damaging such a reflection could be. I think it is tempting to understand the idea of ‘evolution making more likely’ as ‘evolution *exerting some weak-ish control over us*’. But this is probably just a mistake: given the fact that someone put twice as many black balls as white balls into this urn, it’s more likely that I’ll select a black one; but that doesn’t mean that our putting twice as many black balls in as white ones *made* this one black.<sup>12</sup>

Thankfully, however, we can leave the whole issue behind us since, it appears, the etiological skeptical skeptic is not committed to putting her challenge in terms of *evolution*.<sup>13</sup> As David Enoch has recently argued, the real problem for the non-naturalist moral realist is not so much that the story about ‘where we got’ our moral beliefs is disconcerting; the real problem is that this kind of realist appears to be barred from *any* kind of explanation for how we might be *reliable* in our moral judgments—or, in other terms, barred from explaining the ‘coincidence’ (in

---

<sup>11</sup> This point would also have to be kept separate from the very plausible, but surely harmless, thesis that many of our moral beliefs track certain evolutionary facts for the very good reason that certain rules only apply to us in virtue of particular (and no doubt *contingent*) facts about our biology.

<sup>12</sup> Cf. White’s related discussion of “Adam’s Party” at 586ff. There are further possibilities here. It might be said that it makes sense, given our evolutionary history, that we should have a *disposition to believe* some moral systems over others—say, kin-centric ones. The disposition is then non-truth tracking; but explains our beliefs; this is supposed to engender skepticism. Once again, I am not at all sure just what *dispositions to believe* I’d expect evolutionary history to equip us with; and the variety of *actual* moral views might make us doubt that there really *are* any such dispositions in the area. If we could be so convinced, we might take the patent difficulty of inculcating moral beliefs into children to be evidence that ‘our’ moral system in fact runs *contrary* to what evolution has ‘disposed’ us to believe. But evidently a sounder understanding of a *disposition* would be required to sort this out.

<sup>13</sup> Street herself concedes this at p. 155 (2006). The argument of Street, n.d., relies less heavily on evolution, speaking instead of “upbringing, history, culture, education” as well as “genetic inheritance” (e.g., at p.12). In fact, ‘evolution’ plays a rather different role in this argument: it is wheeled in to shore up a disanalogy between normative beliefs and (roughly) perceptual beliefs: see pp. 25ff.

the most anodyne sense of that word) of *my moral beliefs* and *the moral facts*.<sup>14</sup> Whatever the causal story is which explains where our moral beliefs came from, it is hard to see how that story could explain how we might be *reliable* here: for the moral facts are supposed (by the non-naturalist) to be *causally inert*; and the realist can hardly allow that our moral beliefs *compose* the moral facts without giving up on realism.<sup>15</sup>

The worry could be put this way. Although every adult recognizes that not every moral judgment she has ever made was correct, and so recognizes her capacity for error; nevertheless we take it that, generally speaking, we are indeed reliable in moral matters: we would, when presented with some situation for moral evaluation, judge correctly (most of the time). And yet this can seem quite mysterious: just what explains this talent?

Nor does an argument of this shape need to presuppose that we begin without any justified moral beliefs. It can allow that, initially, we have some justification for (say) moral judgment M in the form of other moral propositions. The idea is that, having discovered that the correlation between the moral facts and my moral beliefs is inexplicable, our initial justification is undermined: if it is inexplicable how our moral-belief-forming mechanism could be reliable,

---

<sup>14</sup> See Enoch 2010, esp. p. 421ff.: “So there is a correlation between...normative truths and our normative judgments. What explains this correlation?” Enoch trades back and forth between putting the problem in terms of *reliability* and putting it in terms of the “correlation between...normative truths and our normative judgments.” I come back to this issue in §3 below. Cf. also Field 1989 (p.25ff) for a parallel argument against mathematical Platonism.

<sup>15</sup> That is Street’s answer. Among other possible interpretations of our problem—characterized as the one “raised by the apparent causes of our beliefs”—White considers something like the version just given in the text (attributing it to Schechter) at 592ff. of his 2010. (He also notes that it might be what “lie[s] behind” Street’s argument.) But he dismisses it rather glibly, on the strength of the following analogy: suppose we had a robot which was given a randomly produced program for generating strings of arithmetical symbols; now suppose that, on examining its output, we discovered that the robot was emitting all and only true arithmetical propositions. In that case, White writes, “the tension...seems to push in just one direction: against the assumption that the program was randomly produced.” But as a solution to *our* difficulty, this seems doubly mistaken: what, in the first place, would be the analogue of ‘not randomly produced’? Who or what can we suppose to have given us our non-random program for generating moral beliefs? Then too, it may be relevant here that, though we can (I think) imagine very different, but internally coherent moral systems, it is much harder to do the same for math. This matters, since part of what generated the tension in the original challenge was the fact that we must believe that *our* system is superior to *others*’ (if only *hypothetical* and *imaginable* others); if there is only *one coherent system*, the challenge becomes, if not trivial, anyway much less worrying.

then we are blocked from using any beliefs formed in this way as justification for anything.<sup>16</sup> The upshot is that moral realism appears to entail moral *skepticism*: moral realists are apparently committed to a view according to which there are indeed ‘mind-independent’ moral facts, but it is impossible to explain how we are onto them—we could all be wildly wrong.

The obvious contrast case will be classes of *perceptual* belief. If, for example, our friend Jones claims to know all sorts of facts about the dark side of the moon and, consulting NASA data, we discover that there is, as a matter of fact, a wonderfully high correlation between Jones’s beliefs and the facts-as-reported-by-NASA, then, I think, we will be in the market for some kind of explanation of this correlation. We will be satisfied if we come to believe, say, that Jones was an astronaut, or anyway had *been* to the moon, and *seen* the various things he reports. In short, the explanation for the ‘correlation’ of belief and fact is *a mode of brute access* to the facts, sc., Jones’s faculty of sight. The point is sometimes put<sup>17</sup> in terms of a “causal route” from the facts to Jones, or anyway a causal story about perception; and although this way of putting it strikes me as somewhat misleading, it is useful for reminding us of another way that we could be satisfied with Jones, sc., that he might claim to be in possession of *evidence* for the various facts he reports: perhaps something about the moon’s movements is good evidence that there is a crater on the dark side, or something of the sort. Very roughly, the movements would be *symptoms* of the crater, and hence evidence for it. And now the problem is supposed to be that, insofar as we take moral facts to be ‘causally inert’, it will be impossible to see either how we could have some (causally understood) faculty for directly apprehending moral facts (analogous to sight); or how there could be *evidence* for—in other words, items causally downstream of and

---

<sup>16</sup> Cf. Enoch, pp. 423-4. The requirement that we *first* be able to tell a story about why our moral judgments might be reliable before acquiring any moral justification is too strong; such a requirement would have equally devastating effects for (e.g.) perception.

<sup>17</sup> E.g. by Enoch at p. 421; Street n.d., also invokes “causal forces,” at p. 25 *et alibi*.

hence symptoms of—these causally inert moral facts. The conclusion will be that we must either relinquish non-naturalist realism, or else become skeptics about moral knowledge.

It may now seem that the form of the challenge has nothing particularly to do with the *etiology* of our moral beliefs at all; but I think that would be too hasty a conclusion. I return to this point below.

3. Nevertheless, there is perhaps something dubious about this talk of ‘explaining our reliability’. Enoch’s thought,<sup>18</sup> as we have seen, is that if we have no idea how we might explain our reliability in moral matters, then we either have to settle for an inexplicable correlation (between belief and fact) or, if that is unacceptable, give up on regarding our moral faculty as reliable; and therefore give up on our ‘internal’ justification; and therefore become skeptics about moral knowledge. But what exactly would it mean to be *reliable* in moral matters?

It is important in the first place to recognize that the challenge is not to *certify* or non-circularly *justify* our reliability; if that were the challenge, perception would be in equally bad shape.<sup>19</sup> The challenge is rather to *explain* our reliability.

But reliability in what? If the question were how it is that, over time, I can be expected to make correct moral judgments, then surely the answer is: I have some basic stock of moral beliefs which are true. Since my future judgments are rooted in these basic beliefs, they are likely, though of course not guaranteed, to be correct.<sup>20</sup>

---

<sup>18</sup> Street n.d. is plausibly interpreted this way, too. What she adds is a story which is supposed to be a more plausible explanation for the source of our beliefs (than any realist story). Joshua Stuchlik, in an unpublished MS, takes a similar line, and calls the challenge the ‘Argument from Reliability’.

<sup>19</sup> Cf. White 2010, p. 591; 603-4; see also fn. 16 above.

<sup>20</sup> These beliefs need not be ‘basic’ in any especially demanding sense; the point is that the reliability of (occurrent) moral *judgments* is plausibly explained by some stock of (standing) moral *beliefs*.

Attention will focus, then, on this stock of basic moral beliefs. And the question becomes: how can we explain the correlation between *these* beliefs, and the moral facts? In terms of reliability, the point becomes: we have no way of explaining why it is that this stock of moral beliefs should have been *more likely* to be true than false. Whereas, in the perceptual case, this is just what we do have; for the causal story, no matter how sketchily described, appears to explain why it is that our perceptual beliefs are more likely to be true than false.<sup>21</sup>

4. One way to get a purchase on things here is to consider the other obvious way in which the correlation of moral belief and fact might be explained, sc., a kind of *constitutive* account according to which the moral facts are somehow or other dependent on, or a function of, our moral beliefs. This is the essence of Street's alternative to realism, a view she calls 'constructivism'.<sup>22</sup> The view is, very roughly, that the normative facts—facts about what I have reason to do—are a “complex function” of my various other normative beliefs, evaluative attitudes, desires, and so on. In particular, R is a reason for me to  $\Phi$ , just in case that very judgment (“R is a reason for me to  $\Phi$ ”) withstands scrutiny from the perspective of my other judgments about what I have reason to do.<sup>23</sup> Street characterizes her view as a brand of antirealism; and it is supposed to have, as we saw, the distinct merit of being able to explain the ‘correlation’ between moral belief and fact; as she puts it:

---

<sup>21</sup> This distinction between two kinds of reliability—reliability as ‘likely to get judgments right in new cases’, and reliability as ‘likely to have one’s standing beliefs in this domain be true’—is masked, I think, by the fact that the explanation is *the same* in the perceptual case; whereas they potentially come apart in the ethical case. This might also explain Enoch’s (otherwise very odd) definition of ‘reliability’: “a class of beliefs is reliable...if and only if a sufficiently large portion of it is true” (418). (In fairness, Enoch flags his definition as unusual.)

<sup>22</sup> Street’s view is presented as a *metanormative*, rather than *metaethical* view (see n.d., *passim*), and so I discuss it in those terms; see fn. 4, above.

<sup>23</sup> See Street, n.d., p. 13. Street speaks of “one’s own set of evaluative attitudes,” but this phrase covers a host of desiderative states and evaluative judgments. The story perhaps recalls Williams’s in “Internal and External Reasons,” a point which Street notes at n.d., pp. 4-5, 35.

If constructivism is true, then *of course* there's a striking coincidence between true normative judgments, on the one hand, and the normative judgments that causal forces led us to make, on the other, because normative truth just *is* a (rather complex) function of the normative judgments that causal forces led us to make.<sup>24</sup>

There is an obvious sense in which such a view strikes us as 'antirealist': what I have reason to do is not, ultimately, entirely independent of what I *believe* I have reason to do. But, interestingly, on certain familiar definitions, Street's view is technically a *realist* view after all: it is possible (however improbable) for an agent to make *mistakes* about what he has reason to do—he may, e.g., fail to give his beliefs sufficient 'scrutiny'; fail to weigh his various reasons against each other properly; and so on.<sup>25</sup>

Now no one would deny, I think, that Constructivism is anti-realist in *spirit*, even if it comes out as a form of realism on one familiar scheme; there are, in the last analysis, no facts about what I have reason to do to do that are *completely independent* of what I want, am in favor of, believe to be a reason, etc.—in short, no facts about what I ought to do independent of my other evaluative attitudes and judgments. As we might put it: though there are kinds of mistake that one *can* make, there are other kinds of mistake one *can't* make.

---

<sup>24</sup> See Street, n.d., p. 14. As a matter of fact, it is not so clear that Street's solution solves the puzzle (which, again, she poses as one for *normative* beliefs) in the way that she claims for it. If we accepted Constructivism, we would, it appears, have an explanation for why my normative beliefs about what I have reason to do largely coincide with the facts about what I have reason to do. But this is surely just *one subset* of my normative beliefs: the self-regarding or 'first-personal' ones. And so far from providing an *explanation* for why my 'third-personal' or other-regarding normative beliefs match the (third-personal) moral facts, I shall, having accepted Constructivism, have to concede that I was *wrong* about some (potentially) very large subset of my normative beliefs: for instance, "No one has a reason to torture puppies for fun." It's surely hopeless to try to get a fix on what 'percentage' of my normative beliefs I'll have to give up (that would require getting a fix on the number of total normative beliefs I hold, surely an absurd task); but insofar as there are any normative beliefs I can be said to 'have', a lot of them will be 'third-personal' (other-regarding, really) in the sense of the example just given. And these will all have to go. So we might say: *half* the work of meeting the epistemological challenge is done by the 'constitutivist' move; half the work is done by junking an enormous number of normative beliefs. This consequence of the theory is obscured in Street's exposition by her use of the *plural* first-person pronouns. This point is picked up below, in the text.

<sup>25</sup> The 'familiar' definition of realism is: we are realists about some domain if we regard propositions in that domain as truth-evaluable and not all false. I myself prefer the more minimal 'possible to make mistakes', but nothing here turns on it; indeed, nothing turns on how we classify constructivism, either.

But the fact that Constructivism is, in one sense anyway, a kind of realism, is important if only for this reason: supposing that Street is right that Constructivism meets the epistemological challenge that she sets up, it transpires that there *are* forms of realism that can indeed meet the challenge. This reflection might inspire us to wonder just what *other* kinds of realism might meet the challenge—whether there might not be more *robust*, and in particular non-subjectivist, versions which do. To determine that, it would be useful to look at the broad structure of the puzzle, with the aim of working out an abstract characterization of the kinds of realism that might succeed.

5. Evidently, we need, in the first place, for it to be possible to be *wrong*—this gives us realism. But another requirement for a form of robust realism must be this: it must be possible to locate at least one class of error exactly where people have (apparent) *disagreements* with each other in ethical matters. That is, it must be possible to understand the disagreements between two parties as plausibly turning on a *mistake* that one party is making. We need this element in order to prevent the view from becoming a variety of subjectivism, a point which can be brought out by again contrasting constructivism. We saw that constructivism allows one narrow kind of error—roughly, an accounting error. But although, if constructivism were true, it would be *possible*, when I say to Caligula, “You have a reason to stop torturing that puppy,” and he contradicts me, to suppose that Caligula has made an accounting error—has actually misapplied the relevant function to his other evaluative attitudes; has, in effect, gotten wrong what he *actually* has reason to do—although it would be possible to suppose this, in practice, no one ever would (or does).<sup>26</sup>

---

<sup>26</sup> And yet it may be worth saying that, if we were truly banned from using the language of ‘external reasons’, it might be that this is exactly what we *would* dispute. As Hobbes memorably put it: “For I doubt not, but if it had been a thing contrary to any man’s right of dominion, or the interest of men that have dominion, that the three angles

If, conversely, we *can* understand standard cases of ethical disagreement as turning on an *error* made by one party, then we open the way for an objectivist epistemology.

We need this possibility, in any case, for a *sane* realism. But is *this* really what we need to ward off *subjectivism*? After all, isn't 'ultimately not independent of my attitudes' sufficient for subjectivism? –Obviously there is no point in quarrelling over the word. But consider: what should we call a view that allowed that all facts about *what I ought to do* were a function of (say) my desires; but also allowed that I am *often wrong* about my desires? The possibility of construing things this way is perhaps not so far-fetched: we *do* say sometimes, "You think you want this, but you don't really." My point is not to vindicate such expressions (it will be objected that they are, in some sense, figurative language; or perhaps someone will say that desiderative states are 'transparent'); but only to insist that a sane realism *could* allow that (anyway) facts about *what I ought to do* supervene entirely on facts about *what I want to do*—so long as it allowed that disagreements between Jones and Smith over what Jones ought to do could be construed as disagreements over what Jones desires; and that Jones might indeed be the one making the mistake.

Perhaps such a view should nevertheless be called 'subjectivist'; it seems in any case to be a kind of subjectivism we can live with. Subjectivist theories become *revisionist*—and, by that token, disturbing—only when they insist that Jones cannot be contradicted (say, by Smith) when he makes a judgment about what he has reason to do. What we need, again, is for it to be

---

of a triangle should be equal to two angles of a square, that doctrine should have been, if not disputed, yet by the burning of all books of geometry suppressed, as far as he whom it concerned was able" (*Leviathan*, Part 1, Ch. XI, §21).

possible to construe ethical disagreements as turning on an *error* by one party, where there is a real possibility that *either* party is making the error.<sup>27</sup>

On the other hand, of course, we need this kind of error to be immune to the etiological challenge: it must not be possible to re-raise the skeptical puzzle about the kind of error that our view makes possible.

We have, then, three desiderata for a sound moral realism: the view (1) must allow for the possibility of *error* or *mistake* (that ensures *realism*); (2) that error—or anyway, *some* form of error—must be locatable exactly where moral disagreements break out (that ensures objectivism); and (3) it must not be possible to raise the skeptical worry around the possibility of this kind of error.

Putting things the other way round, we could say that ‘etiological’ skepticisms evidently turn on the possibility of a *certain kind of error*. In one way, this is obvious, since to suppose, as any skepticism must, that we *might be wrong*—wildly wrong, even—about all of our judgments, is to suppose that we might be making *some* kind of *mistake*. But the importance of this point can only emerge when we recognize that there are *different species* of possible error. What we would need, then, to defeat such skepticisms, is to construct something which would count as a realism—something, that is, which would allow for a *certain* kind of ‘robust’ error, along the lines outlined above—but on which an error *of the kind imputed by such skepticism* is not

---

<sup>27</sup> This last reflection points to a possible tension: it may be hard to sustain a ‘revisionist’ subjectivism that doesn’t collapse into antirealism. For, insofar as a subjectivism allows that one can be mistaken (i.e., insofar as it is not a form of antirealism), it will be possible (however implausible) to locate ethical disagreements just where these mistakes are, and therefore for anyone to be *contradicted*; and so the subjectivism will not be revisionist after all. (See fn. 24.) As we could put it: the way in which Jones “can’t be contradicted” must not turn on the metaphysical impossibility of error (that would be antirealism); but only on something like Jones being in a relatively (to everyone else) advantageous epistemic position with respect to reasons-for-Jones. Since *some* version of the last thesis is plausible on *every* metaethical view, it may transpire that subjectivism is a relative affair.

possible. To do *that*, we obviously must begin by isolating the species of error which such skepticisms presuppose. First, then, a detour into *kinds of error*.

6. In order to get a fix on this idea of different ‘kinds of error’, I want to introduce an apparatus for distinguishing kinds of epistemic support.

Now the word ‘evidence’ is often used as a generic term for the epistemic support given by some justifier for some conclusion. I have no quarrel with that use of the word; but in what follows I’d like to distinguish between two quite different ways in which justifiers can support a conclusion; and restrict the word ‘evidence’ to one species of justifier.<sup>28</sup> The possibility of such a distinction turns, as will become clear, on a further distinction, sc., that between (what we could call) *different levels of directness of epistemic access*.

The paradigm case of *evidential* support, as I shall be using the term, might be a (certain kind of) jury trial: on the basis of (what we indeed call) *evidence*, we are supposed to come to a conclusion about whether (say) Jones killed Smith. It seems clear enough how such ‘evidence’ might support that conclusion; roughly, we reason by inference to the best explanation. So, e.g., the best explanation for the footprint on the lawn, the blood in Jones’s car, and so on and so forth, is (let’s suppose) *that Jones killed Smith*. The blood, the footprint—these are understood as *symptoms* of the murder-of-Smith-by-Jones in a familiar sense: a causal chain connecting them to such an event type makes them *reliable indicators* of it. By the same token, a conclusion on the basis of evidence—“Jones killed Smith”—is undermined when we are given an alternative explanation which is more plausible as the *source* of this evidence.

---

<sup>28</sup> I think that this is in line with ordinary usage, but that is irrelevant.

Now sometimes the evidence can be *overwhelming*—only a fool would deny that Jones did it. But for all that, there remains—what we could call—a mode of more direct epistemic access to the facts; here, that mode is plausibly ‘seeing that’. Of course, we are supposing that no one did see Jones kill Smith, or else the footprint and the blood would all be *de trop*.<sup>29</sup> In fact, this is the feature of ‘evidential support’ to which I’d like to draw attention: if we have, or come to have, direct access to the facts—say, the fact that *p*—the evidence drops away as irrelevant; so that, e.g., if I *see* that *p*, it would make no sense to reproach me with evidence that  $\sim p$ —and, more to the point, *ridiculous* to *add* anything about the evidence *for* the proposition that *p*. Having obtained direct access to the fact, the evidence drops away as irrelevant. The point, of course, is not to deny that, sometimes, we might even come to doubt that some putative mode of direct access was operating correctly; of course we are fallible. The point is only that we don’t treat the blood or the footprint as on all fours with *seeing that*; we treat the latter as a *more direct* mode of access to the facts; plausibly, as *the most* direct form of access. (What could be more direct than seeing that something is so?)

By the same token, to speak of something as *evidence* for some proposition is implicitly to suppose that there *is* some mode of *more direct* epistemic access to the facts—presumably, a mode which, for some practical reason or other, is cut off from us (else we’d have no need of evidence). Something counts as *evidence*, then, only in *contrast* to a *more direct* mode of access. Evidence acts to support some conclusion; but, again, given the mode of more direct access, the evidence drops away as irrelevant.

I now want to contrast a rather different way in which reasons might support conclusions. It can be summed up very quickly by saying that, in the case of this kind of support, it is

---

<sup>29</sup> Obviously I am not denying that we can make mistakes about what we *think we saw*; only that, in ordinary circumstances, seeing that things are so is taken to be a mode of more direct—in fact, brute—access to the facts.

impossible to appreciate some conclusion independently of the justifiers for it. There is no mode of *brute* access to the facts—a mode of epistemic access independent of the justifiers supporting that conclusion. On the contrary, we can only see that the conclusion is true *in light of* the justifiers for it; there would be no such thing as *just seeing* that it is true. Since these units of epistemic support do *not* imply a contrast with a more direct mode of access, they cannot be called ‘evidence’; I propose to call them, simply, ‘reasons’, and the kind of support they provide ‘rational support’. Evidential support and rational support are then, in this terminology, two species of the genus *justification*.

Is there any domain of knowledge plausibly structured by the ‘rational support’ model? Obviously, in a moment, I am going to make the case that moral knowledge has this form. But it would be nice to have another. In fact I think there are several; which one the reader finds persuasive will depend on her philosophical persuasion. Here are two examples, both, incidentally, owing to Anscombe.

First, then, we might consider *knowledge in intention*, on one familiar story about what it is to act intentionally.<sup>30</sup> According to this story, when one acts intentionally, one knows what one is doing, in the sense of being able to explain, or give reasons why, one is doing what she is doing. If one cannot answer the question ‘Why?’, she cannot be said to be intentionally acting: there is no knowing *what* one is doing apart from knowing *why* one is doing it.<sup>31</sup> This point could be put by saying that what we say in response to the question ‘Why?’ does not play the role of *evidence* for the conclusion “I am doing such-and-such”; on the contrary, there is no more direct mode of access to *what I am doing*, a mode which might answer that question independently of the justifiers. We could put the point, in another way, by saying: insofar as I

---

<sup>30</sup> The ‘familiar story’ comes from Anscombe’s monograph *Intention*.

<sup>31</sup> Maybe there are odd or unusual cases; the point is about the paradigm case.

know what I'm (intentionally) doing, I know *why* I'm doing it; and insofar as I know why I'm doing it, I know *that* I am.

Perhaps this sort of case will be felt to be metaphysically mysterious. There are in any event other sorts of cases which appear to have this structure. Anscombe, in fact, draws attention to another sort of example: the domain of what we could call 'owes-facts'—facts about who owes what to whom.<sup>32</sup> Anscombe's thought could be summarized in the following way. Given a certain institutional background, some facts are (what she calls) brute, relative to others; some number out of a range of the former descriptions (say *xyz*) must hold for a description (say *A*) of the latter kind to hold; but there is no *guarantee* that, *xyz* being true, *A* will be true. Thus, "The grocer supplied me with potatoes" and "I signed a bill" are brute, relative to "I owe the grocer such-and-such sum of money." But the truth of the former statements don't guarantee the truth of the latter, since there may be some exceptional circumstances which infirm the inference—it could be, as Anscombe notes, that we are on a movie set. Furthermore, there are other descriptions which are brute, relative to the members of *xyz*: thus "He left a quarter of potatoes at my house" (and perhaps some other facts) are brute relative to the description "He supplied me with a quarter of potatoes."

Three things are worth noticing about this structure. First, the descriptions *xyz* are not *evidence* (in my sense) for *A*: there is no mode of coming to see, e.g., that she owes the grocer such-and-such a sum, independently of seeing that she was supplied with potatoes, and so on. Second, *xyz* doesn't just *mean* *A*: that point is secured by the possibility of exceptional

---

<sup>32</sup> See Anscombe's "On Brute Facts," (1958b). The point of the paper is, in effect, to subvert Hume's famous conclusion that there was something illicit about the inference from an 'is' to an 'ought'. Anscombe's strategy was to show that things are exactly parallel with the inference from 'is' to 'owes'. Some have understood the paper to be making a point in general about *institutional facts*, but I am not sure this is the best way to put it. Cf. also Anscombe's, "Modern Moral Philosophy," (1958a) which recapitulates and expands on the material in "On Brute Facts."

circumstances infirming the relation. Finally, there is no reason to suppose that we can adumbrate, in advance, all possible exceptional circumstances; for, as Anscombe notes, “one can theoretically always suppose a further special context for each special context, which puts *it* in a new light”;<sup>33</sup> consequently, there is no reason to suppose we can give a reductive definition of A.

There may be other regions of thought structured by this form of justification; but these two examples will suffice for present purposes.

7. Now as we noted, etiological skepticisms about morality can—and, truly, should—allow that we *do* have some justification for our moral beliefs: just whatever reasons we have in support of those beliefs. But the idea is that this kind of justification is *undermined*, since we are precluded (by the causally-inert status of moral facts) from giving any explanation of the correlation between fact and belief here. My contention now is that this idea presupposes a very particular sort of error—an error possible, in particular, in regions regimented by our evidence model of justification, and nowhere else.

Why, after all, is our initial justification supposed to be undermined? The plausibility of that contention turns, I think, on treating those (putative) facts, which are cast in the role of justification for some conclusion, as *evidence*—that is, as *symptoms* of the fact-to-be-justified. And what is true, I think, is that we have no notion of how anything could be a *symptom* of something non-causal. The conclusion from this premise—that justification here has the form of evidence—must be that the force of this evidence is undermined.

We can see now why etiological accounts of our moral beliefs—evolution, our idiosyncratic upbringing, and so on—appear to be threatening: these explanations for the *source*

---

<sup>33</sup> 1958b, p. 70.

or *origin* of our moral beliefs are intended to play the role of *alternative* or *rival* explanations—alternative to any explanation that invokes ‘moral facts’.<sup>34</sup>

But perhaps it was a misstep to regard this kind of justification as evidence in the first place. For—and this is the point—what etiological skepticisms are in the business of insinuating is that (unfortunately for us) we cannot preclude the possibility of precisely the ‘jury-trial’ sort of mistake: the kind of mistake which, if the foregoing thoughts were sound, presupposes a *more direct* mode of epistemic access; and thus, here, a kind of epistemic access to the facts which is independent of the justifiers for it.

The upshot of these reflections is that we ought to reject a view according to which such a ‘more direct’ mode of access to the moral facts is possible. In one way, this may seem sensible enough—won’t any such view have to wheel in, at some point, some metaphysically dubious ‘moral faculty’? That seems right enough to me, but the rejection of such ‘brute access’, and with it, the evidence model, may have further-reaching implications. For in rejecting this model, we evidently embrace a view according to which *it is impossible to recognize a moral fact independently of the reasons for it*. I shall return to some of the possible implications of this conclusion at the end.<sup>35</sup>

Here let me note one other consequence: in rejecting the evidential model, what we reject is the idea that the *source* of our moral beliefs matters. The skeptical puzzle, it will be recalled, insinuated that our moral beliefs were in worse shape than our perceptual beliefs, where we had

---

<sup>34</sup> And here the argument makes contact with Harman’s famous discussion of moral explanations.

<sup>35</sup> One possible implication which I can only touch on here is for the topic of moral testimony. The implication is roughly that moral testimony cannot transmit moral knowledge, since it cannot transmit the *reasons in light of which* the (putative) moral fact is true; or, if we wish to change idioms, it can transmit moral *knowledge*, but not moral *understanding* (see Hills 2009 for a view in this range). It might be objected that testimony can *also* transmit the *reasons*. But, as it seems to me, even if it does so, what it cannot transmit—or rather, what it is not guaranteed to transmit—is *seeing that M is so in light of* those reasons. If it *does* transmit this, then I see no reason to call the process ‘moral testimony’: surely no one denies that one man can convince another of some moral conclusion; if there is a controversy over the topic of moral testimony, it is not over *this* fact.

at least some kind of explanation for our reliability. But what was the force of such a causal explanation? My suggestion now is that the causal story explains our reliability in the perceptual case because it explains how our beliefs might be connected to the proper source. If we give up on the idea that the source of our ethical beliefs is relevant to their truth, we give up on one kind of *mistake* that might be made in this domain. That there are other sorts of mistakes—and hence moral *truth* and *knowledge*—was the point of the foregoing arguments. (I shall try to say a little more about this below.)

Should we believe that ethical knowledge has this form? Or is it just a convenient way out of our bind? One clue, I think, that *something* like this must be right might be found in a humble, natural-language observation. When we want to query someone's claim about (e.g.) some fact about middle-sized dry goods—"There's a raccoon in the attic"—we ask him, "How do you know?" In ordinary circumstances, we are satisfied by the reply, "I saw it." On the other hand, we would never ask *that* question in order to query anyone's *moral* claims—say, "Pirating music is wrong"—and, correspondingly, there is no mode of 'direct access' to which anyone could advert. On the contrary, we ask, "*Why?*" when we ask for the credentials of a moral belief; and the answer will be in the form of *reasons to believe*. In assessing the correctness of a moral belief, that is, we don't *care* where it 'came from'.

But how then do we 'explain the correlation' between moral beliefs and moral facts? If what I have written above is sound, then we should reject the question. Given certain presuppositions, the causal story about perception does indeed explain why our perceptual beliefs are, *ceteris paribus*, more likely to be correct than not: their (causal) connection to their

objects explains this.<sup>36</sup> What we should admit, I think, is that if we restrict ourselves to stories about the origins of our moral beliefs, these beliefs are *not* more likely to be correct than not. But this should not be troubling: if ‘wrong source’ is not the name for one of the kinds of error in ethics, we should hardly be troubled by the fact that their etiology does not explain their correctness.<sup>37</sup> We cannot miss what it was impossible to have.

8. Nevertheless, even if the foregoing thoughts are sound, only half the work would be done by quarantining the kind of error that must *not* be possible on a sound moral realism. The other essential task is the one noted above: we need a *positive* characterization of the kind of error that *is* supposed to be, not only possible, but at the center of moral disagreements. Can this be done?

Here is one suggestion: we might construe the pattern of moral justification on the model of ‘owes-facts’, discussed above. Thus, any moral claim *M* is justified by, and indeed, seen to be true in light of, some further descriptions, *xyz*. So, e.g., the claim “You ought to return his sword” is justified by (say) the further claims “That sword belongs to him” and “You have used it for the purpose for which you borrowed it” (or some such thing). But of course, some further fact *w* could impair the claim “You ought to return it”—say, something to the effect that its owner would likely harm himself with it if given it back. Nor is it the case, we might add, that we could list in advance *all* of the possible further conditions which would impair the claim.

Whatever else can be said about its plausibility, what this model provides us with is a way of understanding how ethical disagreements between two parties might turn on an *error*

---

<sup>36</sup> On the other hand I should state here that the precise way in which the causal story ‘explains the correlation’ between belief and fact in the perceptual case is not perfectly clear. Light does not, after all, bounce off *the facts*. What we would need, I think, is a story about the metaphysics of facts; and in particular about the relation between *objects* and facts—murky matters indeed.

<sup>37</sup> There are of course *certain* kinds of ‘source’ stories that *would* affect our confidence in our beliefs: moralized stories. (“Your peer group is corrupt.”) But this is evidently beside the point.

made by one of the parties, where the imputation of that error to one's interlocutor does not involve relapsing into the evidence model of justification: the imputed error is just that one's interlocutor is *missing* something and that, once articulated, he would see it and realize his conclusion has been undermined. And it does seem right to say that many moral disagreements really do have this form: we remind someone of some further fact obtaining and, acknowledging it, he comes to see the truth of *our* claim in light of this 'further fact'. And though it could not be claimed, with a straight face, that all ethical disagreements have this form, nevertheless the fact that *some* do is an important point to register; it reminds us that there can be such a thing as rational argument in ethics; and that not all disputes boil down immediately to (apparently) intractable disagreement.

But of course we know that the *difficult* ethical disagreements do *not* appear to have this shape. The difficult ones are those in which our interlocutor does not even acknowledge the relationship between the (putative) reasons and the conclusion. Are we thrown back then, on imputing to our interlocutor an error of the banished kind?

I think we are not. What we appear to be committed to, by the non-evidential model of moral reasoning, is only this: we must believe, even in cases where we despair of convincing our interlocutor of some moral proposition M, that there *is* nevertheless some further fact obtaining which, if we could articulate it, a) would be something that our interlocutor would accept in itself; and b) would get him to see that M is so. We are committed, that is, to the idea that a kind of rational persuasion is always possible.

This is, at best, a rough sketch of a view. But we should still ask: Might this view satisfy our three desiderata outlined above (§5)? It seems clear that the view does allow for *error*; besides basic errors of fact, there is this, rather distinctive kind: it is always possible to suppose,

anyway in the abstract, that not all of the facts are in, that there might be some *further* consideration which, once pointed out, would reverse our conclusion. I have just been arguing that the second condition is met as well: it will always be possible to construe ethical disagreement as *turning* on this kind of error. But the central test is the last one: can the skeptical puzzle be raised here? Are we committed, merely by introducing this kind of error, to an inexplicable correlation of moral *belief* and *fact*?

It seems to me we are not. Recall that the puzzle *allows* that some of our moral beliefs may, initially, be justified. If I was correct above, then this initial justification is undermined by the puzzle only if we regard the *kind* of justification we have too narrowly—if we understand it as *evidence*. For in that case it really will be mysterious how we could have any such evidence. But if our justification for our moral beliefs does *not* have the form of evidence—is *not* to be regarded as something causally downstream of the facts—then, I think, our justification is not undermined; and we are therefore in a position to explain the correlation of belief and fact. –It will be, to be sure, a circular story; but that leaves our moral beliefs no worse off than, say, our perceptual beliefs.<sup>38</sup>

9. So much for moral disagreement. What emerges, for moral epistemology, is a picture according to which we can't be *radically* cut off from the moral facts: even if we are wrong now, there must be a *path* from our current set of beliefs to the correct ones.

---

<sup>38</sup> See fn. 16 above. There may be, ultimately, something dissatisfying about our inability to give an *independent* account of the reliability of some faculty. But this is a familiar issue, and in any case, not local to ethics. Cf. White 603-4.

What we appear to be precluded from doing is supposing that our main way of coming to appreciate moral facts is by way of any kind of evidence.<sup>39</sup> In one way, as I say, this seems sound enough: there is something distinctly off-key about asking for ‘evidence’ that (say) killing the innocent is wrong. (*Evidence* belongs in a box with the enquiry “How do you know?”) But in another way, this may be rather startling: for it has become one of the main methods of normative ethics to treat (what are there called) our ‘moral intuitions’ as a kind of evidence, indeed in the restricted sense of this essay, for moral conclusions; and it supposed that we ought to adopt moral conclusions on the basis of the weight of the evidence. Whereas if the considerations canvassed in this essay are sound, it would appear that adopting *such* a perspective on (what I would prefer to call) our spontaneous moral judgments dooms us to exactly the kind of skepticism which etiological worries threaten to bring down on moral realism.

The problem could be put this way. To treat our moral intuitions as evidence is, in effect, to treat our moral faculties as a kind of black box—intuitions become the output of a machine we do not yet understand. But in that case, the most obvious explanations that suggest themselves for such output will be *causal*, or *etiological* ones—explanations which do not appear to even be in the business of justifying those intuitions.<sup>40</sup> In other words: treating our intuitions as evidence has the effect of casting justifications for those intuitions, on the one hand, and causal explanations, on the other, as *in competition* with each other; but the competition is rigged; the

---

<sup>39</sup> This *can* perhaps happen irregularly: I am not myself very sure of what to say here, but perhaps the fact that *this wise man* believes that M is some evidence that M is true. Following our schema, we must say that, if this is evidence, then there must be a mode of more direct epistemic access to the moral fact; but that will be just—seeing that M is true in light of the reasons for it. There is still no such thing as *brutely* seeing some moral fact to be true. And it would remain true that moral *knowledge* could not be obtained through evidence.

<sup>40</sup> This possibility is exploited selectively by Singer 2005; for criticism, see Berker 2009.

causal story will always emerge as the ‘better explanation’; and the rational support for these intuitions will appear to have been undermined.

But whether or not these last speculations are on the right track, it appears, in any case, that moral realism, in and of itself, does not invite skepticism. If the cost of avoiding such skepticism is indeed giving up on one method of moral theorizing that had, perhaps, little to recommend itself in the first place, then perhaps it is not too high.

## 2 MORAL TESTIMONY?

1. The possibility of ‘moral testimony’ has recently become a topic of interest for moral philosophers. The interest (such as there is) in the subject lies, plausibly, not only in the answers to such questions as “Can there be moral testimony?” and “When should we take it?”, but in the way such answers will ramify into moral epistemology more generally. My own interests lie primarily in the latter; accordingly, my purpose in this paper will be try to get a purchase on some nagging questions in moral epistemology by way of considerations raised by the ‘moral testimony’ topic.

The parties in the debate fall, broadly, into two camps: *Optimists* and *Pessimists*.<sup>41</sup> The terms are, in one way, transparent—Optimists are *for* moral testimony (or sanguine about it) whereas Pessimists are *against* it (or skeptical of it)!—but in fact these terms obscure a mass of subtleties. Thus, intuitively, an Optimist about moral testimony might be someone who believes that *moral knowledge* can be transmitted by testimony; whereas a Pessimist denies this. But as Robert Hopkins<sup>42</sup> uses the term, there is more than one way to be a Pessimist: for one can *allow* that moral knowledge is transmissible by testimony, but deny that it is (in Hopkins’s terms)

---

<sup>41</sup> And, unlike in some debates, there are real, existing philosophers occupying each camp. Jones 1995 and Driver 2006 might be counted Optimists; Hills 2009 is a Pessimist; Hopkins 2007, from whom I borrow the terms, is studiously neutral, but outlines what he takes to be the best strategy for the Pessimists; Nickel 2001 is a cautious Pessimist. Anscombe 1981 is a less clear case. But see below in the text for some clarification and qualification of the terms.

<sup>42</sup> Hopkins 2007, p. 613; as I say, I take the terms over from him.

*usable*; and thus still be a Pessimist.<sup>43</sup> One difficulty (which Hopkins overlooks<sup>44</sup>) with such a terminology is that it seems possible, in fact, to *deny* that moral knowledge is transmitted by testimony, but to *accept* that what *is* transmitted *is* usable.<sup>45</sup> Is such a philosopher an Optimist or a Pessimist? And further complications enter when we ask a Pessimist just why what is transmitted by testimony isn't usable: for some theorists appear to think the difficulty is epistemic or quasi-epistemic; whereas others have held that the difficulty is specifically *moral*: one cannot *act morally* on something that one has acquired from testimony.<sup>46</sup>

Obviously the philosophical matter in the area swings free of any particular labeling scheme; in what follows I shall use the terms in the loose sense in which they were first introduced; but what *is* the philosophical matter? We can identify several, potentially independent questions: Is moral knowledge transmitted by moral testimony? Is what is transmitted by moral testimony 'usable' (in this way or that)? And: If what is transmitted is *not* usable, then why not? In some ways, the first question is the least interesting: we might take a terminological struggle over the word 'knowledge' to be as pointless as one over the words

---

<sup>43</sup> This is, in fact, the strategy that Hopkins suggests that Pessimists adopt: see Hopkins 2009, p. 626 ff. Hopkins introduces a further complication which may be of some importance: a Pessimist, for him, is one who denies "the legitimacy of relying on another's word in moral matters" (p. 613) *unless it is unavoidable* (see, e.g., p. 620-1)—for, Hopkins says, *no one* would deny that there are *any* occasions where we must take another's word in moral matters: his go-to exception is: children learning morality from their parents (see, e.g., p. 612). I say something about this below.

<sup>44</sup> The strategy of Hopkins's paper suggests that he has simply *missed* this possibility: his suggestion, as we shall see, is that the Pessimist should hold that testimony transmits knowledge, but that that knowledge is not usable; the implication is that denying 'usability' is sufficient for Pessimism; is it necessary? Either way we go here seems to deprive the terms (Optimist and Pessimist) of some of their point.

<sup>45</sup> This is close to Jones's position (although it's clear where her heart is): she refrains from claiming that moral testimony transmits knowledge, but insists such testimony is, in any event, usable. She gives a kind of impressionistic argument for the view that knowledge is so transmissible in a footnote, though I do not understand it (see fn. 32, p. 75).

<sup>46</sup> For the view that the difficulty is 'epistemic or quasi-epistemic' see Hills 2009; for the view that the problem stems from a requirement on *acting morally* see, e.g., Nickel 2001.

‘Optimist’ and ‘Pessimist’.<sup>47</sup> The interesting question, it seems to me, is this: what does the testifier *have* that his auditor does not? Having answered *that* question, we shall be in a position, I think, to answer questions about what is transmissible and what is not; and of what is not, why not; and so on.

2. Before turning to that question, one precautionary note is in order: there are *all sorts* of cases here; and they are not always well distinguished by the theorists. Thus the questions about moral testimony are often asked *à propos* of some schema (which then undergoes complications), roughly, “Suppose someone tells someone else some moral proposition M; does the auditor acquire moral knowledge, and if not, what more is needed?” But there are all sorts of things which could be called a ‘moral proposition’; and it may be that very different answers are called for in the different cases. Just to take a few: moral propositions can be *general* (“Theft is wrong”) or *specific* (“*This* is wrong”); can introduce a new moral concept (“This is unchaste”) or attempt to subsume a new case under an old one (“Downloading music from the Internet is theft”); can make reference to the *grounds* for the judgment or be unadorned; and so on.<sup>48</sup> Then too there are questions about whether testimony which introduces something pertinent to a moral judgment, but itself contains no ‘moral content’, is to count as moral testimony<sup>49</sup>; and just what ‘containing moral content’ means.

---

<sup>47</sup> Though it may be that, when the *central* questions are answered, we shall have no trouble deciding what to say. In the end, I offer a qualified defense of the claim that it is precisely moral knowledge that testimony cannot transmit.

<sup>48</sup> Indeed, it seems quite possible, *prima facie*, that the disagreements in the area could be sorted out by more finely distinguishing examples. I don’t actually think that this is so; it would be in any case a rather *boring* philosophical treatment, consisting as it would in a long list of different cases, and the considerations attaching to each. Although it may be that an allergy to the *boring* is an important philosophical handicap.

<sup>49</sup> See Hills 2009, fn. 1 p. 94.

In what follows I shall try my best to be alive to such distinctions. But my thought—or anyway, my pious hope—is that once we get clear on the central question—What does the testifier have that her auditor does not?—the answers to the subtleties will fall into place. Nevertheless, following tradition, my paradigm for a bit of moral testimony will be this: a *particular* claim invoking a thick moral term.<sup>50</sup>

3. Now various reasons have been suggested for why moral testimony might either fail to transmit knowledge, or anyway fail to transmit something that the auditor can *use*: e.g., the *seriousness* of moral decisions; the putative (special) difficulties in identifying moral experts; the requirements of Kantian ‘autonomy’ (sometimes fuzzily classed as a kind of ‘inauthenticity’); the ‘practical’ nature of moral knowledge; the special connection between moral beliefs and character; the non-cognitivist nature of moral judgments; the non-propositional nature of moral judgments; and finally something vaguely described as a requirement that one ‘grasp the reasons’ for some judgment. Some of these suggestions overlap; some cannot be held together with certain others; several are ambiguous, as stated. In general, Optimists have tended to deny only some of these features, and to accept others; and to claim instead that these latter, admittedly special features of ethical judgment pose no special problem for testimony; or that, insofar as they pose a problem, they pose no problem for moral testimony in general, only particular occasions of testimony.<sup>51</sup>

---

<sup>50</sup> Jones 1995’s example is (roughly), “*This is sexism.*” See §3, below. The example is picked up by Driver, Nickel, and Hills.

<sup>51</sup> Thus one, technically Optimistic strategy, has been to concede that moral testimony *very often* fails to transmit something that the auditor can use; but that this is a reflection, not of *in principle* difficulties with moral testimony, but just particular difficulties which happen to arise more often here, but needn’t. Cf. Driver 2006.

Here I think it will be useful to enter a distinction. The ‘special features’ of the moral sphere, just mentioned, that have putative implications for moral testimony, can be classed into two kinds: on the one hand, we have the (supposedly) special features of moral judgment which *cannot be passed on through testimony*, and, not being transmissible, may (or may not) threaten the status of the auditor’s ‘information’ as either *knowledge* or anyway as *usable* (call these ‘transmission-level factors’); and then we have the (putatively) special features of the moral sphere *in general*, which may themselves threaten the status of this information (‘non-transmission-level factors’). So, e.g., if the special ‘weightiness’ or ‘seriousness’ of moral matters is a problem for moral testimony, then this is not (obviously) a problem about *transmission*; whereas if the problem with moral testimony is that it can’t transmit the non-cognitive element of a moral judgment, then this is, evidently, a transmission worry.

Here, then, we encounter the curious fact that nearly *all* parties to the debate—Optimists and Pessimists alike—are united in holding that there is *something* which the testifier has, which the auditor does not, which is *not* transmissible. Thus Karen Jones (an Optimist), employs as her paradigm of moral testimony a case in which a group of women reliably inform a male friend (‘Peter’) about the sexism of some third party, to which sexism Peter is blind.<sup>52</sup> In her story, Peter never comes to see the sexism to which his friends advert; but, Jones contends, Peter ought to *accept* their testimony; and in fact act on it.<sup>53</sup> Driver, another Optimist, also concedes that there is something which the testifier has which his auditor doesn’t, even after testimony; this

---

<sup>52</sup> The name, ‘Peter’, is given by Jones. In this paper I refrain from my usual habit of taking ‘Smith’ and ‘Jones’ as placeholders for names, in the first place because I have to talk about an actual philosopher named ‘Jones’; and in the second because, I am told, one makes himself ridiculous by talking about Smith and Jones. *En passant*: for reasons we need not enter into here, Jones’s example produces in me precisely the opposite intuitions to those she means to elicit.

<sup>53</sup> Jones 1995 suggests, but doesn’t defend the claim, that Peter has hereby acquired moral knowledge; see fn. 45, above.

‘something’ she labels “super-knowledge” and “knowledge-plus.”<sup>54</sup> Interestingly, Driver cashes out the distinction between that knowledge which is passed on by testimony, and that ‘super-knowledge’ which is not, in terms of a distinction between *evidence* for a claim, and (appreciation of the) *reasons* for it. I return to this point below.

This is in one way all very unsurprising: the very idea of moral testimony turns on the idea that the testifier is in some kind of privileged (epistemic?) position. In another way, though, this may seem quite surprising, since holding to such a view, while remaining an Optimist, entails holding, either that one can have moral knowledge without this special ‘something’, or anyway that one can ‘use’ testimony without it; or both.<sup>55</sup>

But a little reflection shows, I think, that this last is not surprising either. For—anticipating—if testimony is to *be* ‘testimony’, there will have to be *something* which is not passed on to the auditor; otherwise, what we should have is, not a case of one person ‘testifying to’ another—or: telling her something—but of one person getting another to *see* something; or getting her to *appreciate* something. We can leave aside for the moment just what these terms (‘see’, ‘appreciate’) come to—indeed, what they come to is apparently (though not necessarily) just whatever the ‘something’ is which, in standard cases of moral testimony, the testifier has and the auditor lacks; so getting clear on the one will have to wait till we get clear on the other.

Now it may be that this way of dividing the terrain will appear to beg the question against the Optimist. But as I have said, the Optimists themselves have tended to focus on cases where *something*—some kind of advantage—possessed by the testifier, isn’t passed on; and I think they have done so—rightly—through an appreciation (implicit or otherwise) that a case in which someone is *gotten to see* something is just not a case of moral *testimony*.

---

<sup>54</sup> Driver 2006, p. 638. These terms are apparently intended to be pejorative.

<sup>55</sup> And this is just the position of Driver and Jones.

For these reasons, in what follows, I mean to set aside, for as long as possible, the questions, *both* of whether or not moral testimony transmits knowledge; and of whether or not what is transmitted is ‘usable’. Since everyone seems to agree that *something* is not passed on, the operative question is ‘what?’ The bulk of this essay will be an attempt to get clear on only this: What, exactly, is *not* passed on by moral testimony? That is: What are the ‘transmission-level factors’?

Before resuming the main thread we might remark another case still, which, though it can be confused with the two cases just mentioned, is in fact rather clearly distinct: the case of moral training. For although theorists of moral testimony sometimes claim that even hard-core Pessimists must accept *some* kind of moral testimony<sup>56</sup>, sc., the kind given by parents to children, the assimilation of this case to ‘testimony’ seems quite wrong.<sup>57</sup> I mean just that it would be a gross distortion to imagine that what parents do, primarily, is merely to ‘tell’ their children various moral rules. On the contrary, children must be *trained* to approve of this and be appalled at that; and no amount of mere *telling* could accomplish this training. Nor is the case of training obviously the same thing as (what I’ve called) getting someone to see something; a fact which can be brought out, I think, at the intuitive level, if we reflect that we sometimes succeed in getting a peer to *see* or *appreciate* some moral fact; but we would not call that ‘training’ her.<sup>58</sup>

Perhaps the impression of having begged the question will have been strengthened, not allayed, by these last points; I mean in any case to support them in what follows. But unless the *prima facie* distinction collapses, I think we should begin, then, by recognizing at least three different cases: mere testimony; getting someone to *see* something; and training.

---

<sup>56</sup> Hopkins 2007, e.g., at p. 612; Hills 2009, p. 94.

<sup>57</sup> And indeed, if this bachelor may be so bold, would, if accepted, conduce to a rather bad style of parenting.

<sup>58</sup> Thus Anscombe has sometimes been taken for an Optimist; but the subject of her 1981 appears to be moral *training*, not testimony.

4. Now as I say, the Optimists largely concede that *something* is not passed on; and this brings them rather close to the apparent Pessimists (and their would-be defenders). Thus Robert Hopkins advises Pessimists to abandon the strategy of denying that moral testimony transmits moral knowledge, and to claim, instead, that such knowledge is not ‘usable’—and unusable precisely because of what is not passed on. In particular, he suggests the following, plausible, if rather vague, “Requirement”: “having the right to a moral belief requires one to grasp the *moral* grounds.”<sup>59</sup>

The difficult question, which Hopkins does not address, is just how to unpack ‘having the right’ and ‘grasp the moral grounds’.<sup>60</sup> These phrases are hardly ‘metaphors’ (dead metaphors, perhaps); but in the present context, it is not quite clear what they come to. An understanding of the former is essential if we are to discover just what it would be *wrong to do* with this (putative) moral knowledge: may I, e.g., pass the testimony on to someone else? And if not, what is the penalty I shall incur in doing so?

But the second phrase—‘grasp the moral grounds’—is, if anything, more essential. Indeed, it is our quarry. Of course, we are all familiar with the (quasi-metaphorical) use of the expression ‘to grasp...’; but difficulties rush in as soon as we give the obvious answers.

Thus Hopkins deploys his Requirement as the upshot of a discussion in which having *moral* grounds for some moral proposition is contrasted with having ‘Humean’ grounds for the truth of that proposition. Humean grounds, for Hopkins, include such considerations as ‘that everyone believes m’; ‘that this guy who’s telling me m is generally reliable’; ‘that he is

---

<sup>59</sup> Hopkins 2007, p. 630, emphasis original.

<sup>60</sup> I mean no criticism of Hopkins: he forwards his Requirement at the end as a gift for future Pessimists, to be worked out by them.

presently being sincere'; and so on; and Hopkins characterizes 'Humean' considerations as "reasons, available to the recipient of testimony, for thinking that what she is told is true."<sup>61</sup>

This strikes me as a rather vague formulation. After all, if the recipient of the testimony is told, in addition to the (putative) fact that  $\Phi$ ing is wrong, the further fact that, "It is wrong because of R," where R is some moral reason, then, by the letter of this definition, R should be counted among the Humean reasons; but this evidently destroys the contrast we wanted to understand.<sup>62</sup> Now we might think to get a purchase on the distinction—between Humean grounds and moral grounds—from the *other* side: Humean grounds for a moral distinction will then just be all admissible grounds that are *not* moral grounds. I actually think that this will not do either, for reasons which will, I hope, emerge below. Suffice it to say for now that the relevant distinction in the area, as I see it, also holds between 'Humean grounds'—that is, what I should like to call 'evidence'—and certain *non*-moral reasons—'grounds', in my terminology. We can get some idea of the distinction, I think, if we note that the 'Humean' method of reasoning outlined by Hopkins is, in effect, a form of inference to the best explanation; whereas—as I shall try to show below—it would be a serious mistake to think that enumerating (e.g.) moral grounds for a moral 'proposition' constituted an exercise in inference to the best explanation. My preferred way of putting the point is this: X is '*evidence*' for Y if and only if there is some *more direct mode of access* to the fact (Y), which mode is *currently* unavailable; where '*grounds*' for Y presuppose no such thing. Then too, I think we can get some sense of the distinction by asking ourselves this question: In which case, when we mean to query someone's

---

<sup>61</sup> *Ibid*, p. 627.

<sup>62</sup> This possibility—of being *told* the moral reasons by the testifier—just appears to be off Hopkins's radar, a point I return to below. In saying this I again do not mean to disparage his careful, patient paper: his interest is just elsewhere.

claim, do we ask, “How do you know?” and in which cases do we ask, “Why?” But I hope to make this clearer in what follows.

At the moment, the important point is just this: if Hopkins’s suggestion is right, then what the testifier has, and his auditor lacks, even after the testimony has been given, is a *grasp of the moral grounds*.<sup>63</sup> But we should immediately add that this ‘grasp’, whatever it is, cannot be passed on, from testifier to auditor, merely by *more testimony*: that is, someone’s telling me (in addition to some moral fact) ‘the *moral grounds*’ for some moral proposition (“Φing is wrong”), does not suffice to give me a ‘grasp’ of those reasons.<sup>64</sup> This comes out, I think, if we reflect on the intuitive fact that *merely being able to rehearse* the grounds for some claim is just not the same thing as ‘grasping’ them. Thus—anticipating—to take an aesthetic example, if I am told (say) that “This symphony is a masterpiece” on the grounds (in part) that it “employs a dramatic use of chromatics in the bass,” I have evidently been given the (aesthetic) proposition and its (aesthetic) grounds; I might even go so far as to (fraudulently?) repeat this to someone else; but it seems clear that my ability to rehearse such facts is no guarantee that I actually *grasp the grounds*, in some intuitive sense of that phrase. Of course, we need more than an intuitive sense; in what follows I try to accomplish that. The present point is just that ‘being in possession of’ the grounds for some claim is, intuitively, not sufficient for *grasping* them.

---

<sup>63</sup> The implicit identification here—between what the auditor lacks and what the testifier has in virtue of which he can testify—is licit, given Hopkins’s formulation: surely what’s included in ‘has a right to’ is ‘may testify to the effect that’.

<sup>64</sup> I put ‘the moral grounds’ in scare-quotes because I myself am skeptical that there is some canonical set of propositions (reasons, whatever) which are *the* grounds for some moral proposition. Some of the reasons for my suspicion will come out below, but I try to argue for this claim, in different ways, in Chapter 3.

5. Why not? Alison Hills tries to answer this question in a recent Pessimistic discussion of moral testimony.<sup>65</sup> Hills (like Hopkins) is willing to concede that moral *knowledge* may be transmitted by testimony, but she insists that something else—sc., moral *understanding*—is not.<sup>66</sup>

Now Hills is sensitive to the fact, just mentioned, that, whatever ‘grasping’ or ‘appreciating’ some moral reason is, it must be distinguished from (what I have called) merely being able to *rehearse* the reasons—from merely ‘believing’ them, as Hills says.<sup>67</sup> One needs, further, “a grasp of the relation between a moral proposition and the reasons why it’s true.” This seems right, but is hardly an advance: the mysterious ‘grasp’ shows up here again. We need some further analysis. Hills’s own account goes by way of *counterfactual dependence*; and in spite of my general sympathy for Hills’s view, I don’t think the account works.

Thus Hills claims that, if (say) Judith merely *tells* Claire that *m* (“Φing is wrong”) *because* of R, then Claire’s belief that Φing is wrong counterfactually depends, not on R—though Claire, *ex hypothesi*, *believes* it—but rather on “what Judith says.”<sup>68</sup> Intuitively, this is not correct: we needn’t choose between the two. Thus if, e.g., Claire was set to murder her neighbor for his car, and refrains after being told, by Judith, that “It would be wrong to kill him”

---

<sup>65</sup> Hills 2009.

<sup>66</sup> ‘Moral understanding’, then, plays the role for Hills that ‘grasping the moral grounds for’ plays for Hopkins—indeed, Hills appears to identify the two at various points in her paper: see, e.g., p. 100; p. 101, where Hills speaks of ‘appreciating the reason why’ rather than ‘grasping’; p. 127; etc.; see her fn. 8, p. 98, for an acknowledgement of the kinship between her views and Hopkins’s.

Nevertheless, Hills attempts to elucidate ‘moral understanding’ by way of a list of *abilities* which are individually necessary and jointly sufficient (or something like it: see her fn. 19, p. 103: “These abilities are, I think, individually necessary for moral understanding, and I suspect they may be jointly sufficient....”). The relevant features include the ability to: follow an explanation for some moral proposition, *m*; explain why *m* in one’s own words; draw the conclusion *m* from the reason, *R*; draw *similar* conclusions on the basis of *similar* grounds; etc. (See p. 102.) But she leaves it open whether or not these abilities are, together, *constitutive* of (‘identical with’) moral understanding, or merely *symptoms* of it (see fn. 18, p. 102). I am broadly sympathetic to Hills’s view; the present paper can be taken as an attempt to rectify some (putative) missteps in her argument; and to suggest that these abilities are, in fact, reflections of something we can characterize more directly.

<sup>67</sup> ‘Believing’ the reasons, for Hills, means: having been told them by someone one finds reliable; sincere; etc. That is: ‘believing’ the reasons for the same ‘Humean’ reasons that one might ‘believe’ testimony about the original moral fact. Whether this deserves the name ‘belief’ might be disputed, but it hardly matters here.

<sup>68</sup> See p. 111, fn. 29; “what Judith says” is evidently to be taken in a *de dicto* sense.

because (say) “It’s wrong to take the life of an innocent”; and furthermore Claire ‘believes’ both propositions (say, on the basis of her estimation of Judith’s ‘reliability’); then there is clearly *some* sense in which Claire’s belief (“It would be wrong to kill him”) is counterfactually dependent on her belief that “It’s wrong to take the life of an innocent”—e.g., we should be inclined to say this, I take it, just in case Claire would have *gone ahead* with the murder *if* she had learned that her neighbor was *not* ‘innocent’. Of course it’s no doubt true that Claire’s *having* the belief—that ‘it’s wrong to take the life of an innocent’—is further dependent on *Judith’s having told her*; but to suppose that the latter was, therefore, the ‘true’ or ‘real’ explanation of Claire’s action, seems no more justified than to suppose that the ‘true’ explanation of Claire’s action was, in fact, Judith’s having been born, since, *that* having failed to have happened, Claire would not have acted as she did.

The last argument is rather compressed. The general point is this: it will not in general do to say, of someone who ostensibly  $\Phi$ s for reason R, that R was *not* her reason, just because she would *not* have *believed* R if something else hadn’t happened. This can be shown in another way, when we consider that ‘learning to *appreciate* that *it would be wrong to do such-and-such* on the grounds that such-and-such-else’ is, as far as Hills’s proposed counterfactual test goes, in just the same position as ‘coming to *believe* that *m*, and that R is the reason for *m*’. And these were precisely the two sorts of situations we had wanted to distinguish.<sup>69</sup> A counterfactual analysis does not appear to cut at the right joints.<sup>70</sup>

---

<sup>69</sup> My own view is that explanations of this kind are infected by a contextual, pragmatic element; so that just what counts as ‘the reason she  $\Phi$ ed’ depends on contextual factors: roughly, what elements of the situation are being held constant by one’s interlocutor; see Chapter 3.

<sup>70</sup> A similar problem affects Hills’s example of ‘Ron’ on pp. 114-15. Adapting a case from Nomy Arpaly, Hills presents the following example (which sounds distinctly like the set-up for a joke): Ron is considering killing Tamara; he asks his rabbi about it, and his rabbi tells him that he oughtn’t do it, since Tamara *is a person*. Ron knows that his rabbi is reliable and trustworthy; therefore, Ron comes to believe that he oughtn’t kill Tamara, and that the reason he oughtn’t is that Tamara is a person. Now, this is just the sort of situation that Hills—and I—want

Then just what *is* involved in ‘grasping the grounds’? I want, in fact, to postpone discussion of just what this comes to until after exploring an alternate suggestion for what is required for moral understanding. The suggestion is, roughly, that the *basic* requirement in the area—in one sense of ‘basic’—is the ability to *recognize* an action (or situation, or etc.) as falling under some moral predicate. And that, though ‘grasping the reasons’ is also essential, it is, in one way, dependent on and logically posterior to the requirement just mentioned.

6. Aesthetic analogies are often useful when discussing ethics, since, it is widely recognized, there are a number of parallels between the two domains. On the other hand, just these parallels entail strict limits on the dialectical usefulness of such analogies, insofar as intuitions about both kinds of case will often align.<sup>71</sup>

Nevertheless, I think the following sort of case may be helpful to consider, in part because of the *disanalogies* between ethics and aesthetics. Take, then, the case of ‘movie judgments’. Just what makes a good film is a complicated, and no doubt controverted thing; still, we are probably all familiar with such judgments, having made many ourselves. So,

---

to distinguish from the case of genuine ‘moral understanding’; and Hills attempts to explain why this doesn’t count as truly ‘appreciating the reason’ by way of a counterfactual analysis; she writes: “Ron’s belief [that it would be wrong to kill Tamara] is...counterfactually dependent on the rabbi’s testimony...together with his [Ron’s] belief that the rabbi is trustworthy and reliable, but not counterfactually dependent on his [Ron’s] belief that Tamara is a person” (see p. 116, fn. 38). But as it seems to me, in order to elicit the intuition that Ron’s belief is indeed counterfactually dependent on “his belief that Tamara is a person,” all we have to do is consider what Ron would do if he came to believe (*mirabile dictu*) that Tamara was *not* a person; presumably, he would kill her. (I.e., he would come to believe, “It would *not* be wrong to kill Tamara.”) The only way, I take it, to block this intuition is to insist, as Hills does later in the same footnote, that Ron would have believed something different (about whether or not he could kill Tamara) if he had been told something different by his rabbi. True, no doubt. But just as Ron’s belief (“It would be wrong to kill Tamara”) depends on ‘what the rabbi told him’, what the rabbi told him depends on further things—say, what is written in the Law. Shall we say, then, that ‘what the rabbi told him’ isn’t part of the relevant explanation either? (In general, we cannot insist that X is not counterfactually dependent on Y, just because Y is, in turn, counterfactually dependent on Z.)

<sup>71</sup> Driver 2006, interestingly, also considers the analogy between ethics and aesthetics; her conclusion is, roughly, that the same general considerations do indeed apply to both domains; but that the special reluctance we have in accepting moral testimony is to be chalked up to a) the greater *seriousness* of moral matters and b) the relative (to aesthetics) difficulty of becoming a moral expert (i.e., to non-transmission-level factors); nevertheless, there is nothing in *principle* wrong with moral testimony. For what it’s worth, (b) seems exactly backwards to me.

consider the case in which I see some movie with a friend; afterwards, I sing its praises, whereas he found it (say) thoroughly insipid. We can imagine the kinds of things he will say if he wants to convince me of its demerits. Now many of these things he may have *noticed*—in the sense of: become consciously aware of—while watching the movie; but others he need not have. Indeed, it is quite common—I find—when we’ve disliked (or liked) some film, to find ourselves hard-pressed to explain just what made it bad (or good).<sup>72</sup> My friend’s initial reaction may have been, for example, wincing at some scene; only later, when trying to convince me, does he manage to articulate his complaint: the scene was (say) *sentimentalist*, in the pejorative sense of that term.<sup>73</sup>

The first point I should like to underline, then, is this: that it may be quite difficult to articulate one’s reasons; my friend might, in fact, fail entirely to articulate them; and yet for all that, he might be quite right in his judgment. Now, according to one plausible line of thought, if someone is *never* able to articulate the reasons for his judgments, such judgments do not amount to *aesthetic judgments* at all—indeed, they do not amount to *judgments*. Whether or not this is so can be dropped for the moment. If we imagine my friend to be, *generally*, able to articulate the reasons for his judgment, then his inability here will hardly jeopardize the status of his current judgment.<sup>74</sup>

But this—familiar—inability, or anyway difficulty, in articulating one’s reasons opens the way for a further important point about aesthetic judgment, a point which can be brought into relief by considering a *bad* picture of such judgments. Thus it is sometimes thought—by

---

<sup>72</sup> In fact, the case of trying to convince someone that some movie, or book, or whatever, was *good*, is notoriously more difficult than trying to persuade someone that it was *bad*, which is one reason why I’ve chosen this form for the analogy. Just why this is, is an interesting question—connected, perhaps, to the old wisdom, *bonum ex integra causa, malum ex quocumque defectu*. One thing we may do here, though, is try to *defuse criticisms*

<sup>73</sup> Cf. McDowell 1979.

<sup>74</sup> I myself am not sure I buy this ‘familiar line of thought’. There may be other ways to distinguish ‘mere reaction’ (or whatever) from ‘genuine aesthetic judgment’: for example, by way of the notion of ‘tutored’ and ‘untutored’ judgments; or ‘well-trained’ vs. ‘untrained’ vs. ‘warped’ sensibilities.

philistines, I suppose—that what the *good critic* does is, in effect, test the film against some set of criteria for *being a good movie*; and thus what *being a good critic* consists in is *knowing what's on this list*. The critic then (on this view) may even ignore his own ‘spontaneous judgment’, deferring, as he must, to the list. That, as I say, is the philistine’s view (I have heard it more or less expressed); and the philistine for this reason rejects the authority of the critic as basically fraudulent. Presented with the critic’s judgment, and the reasons for that judgment—i.e., a deployment of some of the items on the list—the philistine responds, “What do *I* care about your list? I like what I like.”

Not that everyone with such a picture of aesthetic judgment might respond this way; the *snob*, we might say, is one who shares this view of aesthetic criticism with the philistine, but actually *embraces* the critic’s judgment: all *he* wants is to find out what’s on the list—to join the club.<sup>75</sup>

Now no doubt there are many fraudulent critics. But if there is such a thing as a good one, then the difference must be something like this: what comes *first* for him, is the judgment of the movie’s quality; he then articulates why: these are his reasons. If these reasons occur over and over—as plausibly they will—then they could indeed be put on a ‘list’ of criteria. (If they can be codified—i.e., if they can be put in the form of unqualified ‘if...then’ statements—then a system of ‘aesthetic rules’ is possible. This seems distinctly less plausible.) But what seems essential is that he *recognize the movie* as good; and that this recognition *in some sense* precedes the reasons he gives.<sup>76</sup>

---

<sup>75</sup> This illustration of a defective picture of aesthetic judgment was impressed on me by Jack Fortune.

<sup>76</sup> This picture is impossibly crude, but the relevant qualifications would take us too far afield. One such qualification would be that the critic’s judgments would surely be much more determinate than ‘good’: it may be, e.g., ‘a moving drama’, or whatever.

In what sense? Well, at least in this sense, evidently, if the philistine really is wrong: the reasons the critic gives must be *attempts to articulate his spontaneous judgment*.<sup>77</sup> For, if our intuition above was sound, then merely rehearsing the reasons—e.g., in the *absence* of a genuine spontaneous judgment to the effect that the film was good—will evince a kind of fraudulence—or in any case, will be a kind of ‘mere parroting’. So, *one* condition on the rehearsal of some set of reasons counting as ‘an attempt to articulate one’s judgment’ is, naturally enough, that one have made such a judgment. There may be others.

Does the reverse hold? Is there, that is, a kind of logical dependence of *forming a spontaneous (genuine aesthetic) judgment* on *one’s giving (or: being able to give) reasons* for that judgment? As I noted, there is a familiar line of thought according to which a *general* inability to give reasons for one’s judgment would undercut the entitlement to call such things ‘judgments’: they would be, in effect, a kind of exclamation of pleasure.<sup>78</sup> But even if we endorse this line, it seems clear that the inability to articulate one’s reasons in a *particular instance* does not, by itself, undermine the status of that judgment. So: if *forming a spontaneous judgment* and *being able to articulate one’s reasons* are both conditions on ‘having aesthetic understanding’ (as we can now say), then nevertheless there is an asymmetry between the two conditions.

7. Let’s return to our story. What I want to suggest now, on the strength of the foregoing, is that, for my friend’s deprecations of the film to be expressions of—what we can call, with no prejudice to the question of how such ‘understanding’ might be *used*—aesthetic understanding, and not merely some affected snobbery, what he says must be expressions of—in fact, attempts

---

<sup>77</sup> And whether or not they constitute such an attempt should be a logical matter, not a psychological one.

<sup>78</sup> Cf. Wittgenstein, “Lectures on Aesthetics,” esp. §17, in Barrett 1967.

to articulate—a spontaneous judgment. In the absence of such a spontaneous judgment, the film may still be sentimentalist trash, but he doesn't understand that (or why).<sup>79</sup> To say this is of course not to deny that his words may nevertheless have the effect of spoiling the film for me—indeed, they may do so even if the movie was, as a matter of fact, quite good.<sup>80</sup>

Supposing though that he is no fraud, our proxy question will be: What does he give me when he gives me his reasons why the movie was bad? In the best case, he gets me to see what he sees. But if the foregoing thoughts were sound, this doesn't mean that he gets me to (say) 'see' that *swelling strings* were played over the funeral scene (a fact I may indeed have missed), and to accept that *that* kind of thing is sentimentalist. What he's trying to do is to get me to see the movie (or that part of it) *as* sentimentalist *in virtue of* the swelling strings; i.e., he's trying to get me to share his judgment, and his tools are the various features. A little lyrically, we might say: he's trying to get me to stand in the right place, in order to see the thing in the right light.<sup>81</sup>

**8.** Now in the first place, if the analogy is sound, then our tentative conclusion will be that, in the ethical case, what fails to be transmitted by *mere* testimony is just this: a spontaneous (ethical) judgment. But here we should enter another qualification, since it will not always be obvious whether or not someone's utterance really does constitute a 'spontaneous moral judgment'. One temptation here is to appeal some introspectible quality of the spontaneous judgment, but I think that should be resisted: following tradition, I think we should say that my spontaneous moral

---

<sup>79</sup> None of this is to deny that it may be necessary, on the way to earning genuine aesthetic sensitivity, that one should first spend some time practicing, as it were, *getting into* the judgment. Just why one should be inclined to do this is a further, delicate question

<sup>80</sup> In general, it's a lot easier to criticize than it is to appreciate.

<sup>81</sup> But here too it's worth saying that standing in *some* place can probably destroy the appreciation of *any* aesthetic object: seeing early drafts of the novel; finding out biographical details of the artist; and so on. There may always be *some* position from which we can (as it were) see the seams.

judgment is indeed a genuine one just in case I am able, in future cases, to make similar judgments—to ‘go on’ with (say) the relevant thick moral term.<sup>82</sup>

And though nothing I have said so far *demonstrates* it, it seems to me that, insofar as we think that the *fraudulence*, exhibited by someone who (as we can vaguely say) ‘makes use of’ an aesthetic judgment he really does not understand, is exactly parallel to the ‘inauthenticity’, or whatever we choose to call it, in the case of someone making use of a *moral* judgment he does not understand, then I think we shall further think that it is *this*, and not, e.g., the *weightiness* of moral judgments, or the difficulty in finding specifically *moral* experts, that is driving our discomfort with moral testimony. Though this is not to say that the native seriousness of moral judgments doesn’t have *some* important consequences; and here we see the importance of understanding just what ‘having a right to’ comes to (see §4, above). So, for example, one way I might ‘make use of’ someone’s aesthetic judgment (about a film, say), is to go see the movie myself; and obviously there’s nothing fraudulent here. Just as surely, there are cases where we blamelessly take another’s word on moral matters: if someone tells me that (say) sitting in this seat (on the 71A) is *terribly wrong* or *deeply unfair*, and I don’t see it at all, surely it would be a little hysterical to call my moving to another seat three feet over ‘inauthentic’ or otherwise troubling; or even, I think, to insist that I would be ‘within my rights’ to keep my seat: when the consequences are trivial enough, the moral claim serious enough, and the testifier basically trustworthy, it seems plausible to say that I ought to just ‘make use of’ the testimony, and act.

Perhaps this will be disputed; it is in any case not my main concern. My concern, at the moment, is this question: is this phenomenon local to these two kinds of case, moral and

---

<sup>82</sup> Cf. McDowell 1979, 1981. This chimes, too, with Hills’s suggestion that one criterion for ‘moral understanding’ is drawing similar conclusions in similar cases (see p. 102); see also her Wittgensteinian-themed footnote 19, p. 103: she too disparages the idea that the achievement of moral understanding can be recognized by “certain sorts of feeling: a flash of enlightenment; a light dawning.”

aesthetic? I noted before that the two spheres are often classed together—plausibly, since, whatever else is involved, *sentiment* or *feeling* is evidently at play in both aesthetic and moral judgments. (Or anyway it has been widely supposed.) On the other side, we have cases of ‘normal empirical judgments’, like, “She’s in the next room,” which appear to pose no parallel difficulties for testimony. Is this the relevant line then, with empirical judgments on one side, and ethical and aesthetic judgments on the other?

I am not so sure that it is. Imagine a man who was born and raised on a desert island; somehow, he speaks English. Upon his being brought to civilization, we attempt to teach him the names of various artifacts, naturally not found on desert islands; but he has some trouble with the concept ‘chair’. Our attempts at getting him to recognize them are in vain. Perhaps I point out to him a chair, identify it. He asks, “Why is that a *chair*?” I give him the stock reasons: arms, back, seat, etc. And yet he cannot consistently use the term; he calls all sorts of non-chair things ‘chair’ and vice versa. (Perhaps at some point he tries to sit on my lap: “But you have arms, back, and a seat!” he complains.) Of course, any particular chair, having been identified by him, will pose no special problems: he will be able to say of *it* that it is a chair; he may even go on to rehearse the reasons (say, to some third party), perhaps quite convincingly: “Oh, you know, it has arms, a back, a seat.” Shall we say he *knows* that it’s a chair (on the basis of my testimony)? It again seems to me entirely indifferent how we answer that question; perhaps he can even ‘use’ the information (someone asks for a chair, and he brings it). But it does seem to me that what he lacks, in this case, is parallel to what someone lacks in the cases of moral and aesthetic testimony. –Parallel, but not the same, since, for one thing, this man’s deficit is not just

an inability to apply the concept ('chair') in particular cases, but a failure to grasp the concept altogether. What he lacks is the ability to make genuine, spontaneous chair-judgments.<sup>83</sup>

A man who doesn't know what chairs are, and has a conspicuously hard time learning it, is perhaps a little hard to imagine. I use the example because it brings to light what I think is an important feature of these cases, sc., our inability to specify individually necessary and jointly sufficient conditions for the application of the concept. For it is a notorious fact that doing so for the concept 'chair' is impossible. This last feature is a little less clear (since I hardly know how to use the terms myself) in the following example, which is in other ways easier to imagine: someone who (say) can't tell the difference between a *sofa* and a *settee*. Here too we can imagine someone who can *profit* by being told (e.g.) that *this* is a *sofa*; but to whom we should hardly award control of the concept, insofar as he can't tell the two apart, quite generally. He cannot, again, make spontaneous sofa-judgments.

On the other hand, insofar as the reasons, or grounds, for some judgment *do* constitute necessary and sufficient conditions for that judgment—and these grounds, or 'features', are themselves understood by the auditor—then I think our worries about testimony evaporate (here). For if there is (e.g.) some one feature that distinguishes sofas and settees (I have no idea myself), then I think that we *shall* say that he learns what a (say) sofa is by being given the reason. Here, 'being in possession of' the grounds and 'grasping them' are the same.<sup>84 85</sup>

---

<sup>83</sup> And note that there is something distinctly off-key about the phrase "*evidence* that this is a chair," said of someone who's (say) looking at a chair. I should note here, too, that though, in this example, our man lacks the concept 'chair' altogether, the point carries over, so far as I can see, to situations where someone merely doesn't know how to apply the term in some more limited class of cases. See below in the text.

<sup>84</sup> Of course, the auditor will have to be able to reliably identify the *feature*; and, in this example, anyway, be able to (as we could put it) use the disjunctive predicate 'is a sofa or a settee'. I don't think these complications affect the main argument.

<sup>85</sup> But here I am decidedly less sure. The difficulty is this: there is some kind of intuitive introspective difference—perhaps illusory for all that—between my seeing, on the basis of some individually necessary and jointly sufficient criteria, that something is an X; and my seeing *immediately* that something is an X, though the conditions for being

9. If this is right, then *one* reason that moral testimony—“That was sexist”—is insufficient for transmitting genuine moral understanding is that such testimony—even when accompanied by further testimony to the effect that this and that are the *reasons* for the judgment—does not guarantee that one can go on using the term; and this in turn, I want to suggest, is because such terms cannot be given necessary and sufficient conditions for their application. That is: whatever reasons we give for the judgment, they only entail their conclusion (the judgment itself) when supplied with a *ceteris paribus* clause. Or again: the judgment goes ‘beyond’ any of the reasons we could enumerate; and so (merely) relating those reasons does not guarantee that the auditor has picked up the concept.<sup>86</sup>

None of this is to say that all cases of moral testimony involve an auditor who has little or no grasp of a concept; indeed, I noted at the outset that there are all kinds of cases; and in a longer discussion, they too would be treated, and perhaps quite differently. But all the cases imagined by the theorists either are<sup>87</sup> or can be massaged into<sup>88</sup> cases where what the auditor lacks is the ability to apply some thick moral term either altogether, or anyway in some class of cases.

Is this the *only* reason that moral testimony is insufficient to transmit moral understanding? I’m not sure myself just what to say here. Intuitively, being able to apply and reapply some concept is *not* sufficient for ‘moral understanding’ since—it seems—what is required in addition is at least that I share (roughly) the *attitude* of the testifier. If, for example,

---

an X are in this way enumerable, and moreover, open to my view. But whatever ‘non-inferential’ knowledge is, it should not be something which is distinguished by *introspection*.

<sup>86</sup> This is my proposed substitute, in effect, for Hills’s counterfactual analysis. My thought is the ‘ability to go on’ is one condition on some apparent spontaneous moral judgment in fact *being* such a judgment.

<sup>87</sup> Jones’s example of Peter and sexism (adopted by Driver, Hills, and Hopkins); Hills’s discussion of lying; etc.

<sup>88</sup> Driver’s story of ‘unjustified killing’; Hopkins’s discussion of workers contemplating a strike; Nickel’s case of killing in self-defense; etc.

Peter is presented with all sorts of cases, and he reliably discriminates between cases of sexism and cases of (say) reasonable discrimination on the basis of sex, nevertheless it seems rather odd to say that he's acquired the concept if he doesn't see the cases of sexism as (in some sense) *bad*. Whether such a thing is even possible is not entirely obvious.<sup>89</sup> In the first place we should have to get clear on just what 'see as bad' meant. I suppose, at a minimum, this would mean that Peter tries himself not to behave in these ways; and discourages it in others; etc. Could Peter reliably make the judgments but not do these things? This seems dimly imaginable; but it does seem right to deny that Peter, in this case, was making the relevant spontaneous judgments.

But in any case a decision here can be left in abeyance since, for my purposes, what is important is that the ability to use the term in novel cases should be a *necessary* condition on making a genuine moral judgment. And that this is so is shown, I think—if the foregoing has been accepted—by the fact that if (what's again only dimly imaginable) Peter acquires, by way of a bit of testimony, the relevant *attitude* towards some instance of sexist behavior, it nevertheless seems rather wrong to say of him that his judgment amounts to a genuine judgment that something is sexist if he cannot identify new examples of sexism in the future.

My view, then, in effect, is this: what's basic to moral understanding is having a spontaneous moral judgment about some situation, real or imagined, in sight or described. But one of the conditions on some judgment's *being* a genuine moral judgment is that I can 'go on'; and this is most clearly seen in the case of command of some thick moral term. It may be that a further condition is that I share the 'attitude' of the testifier; indeed, this is plausible. But 'being able to give the reasons' for one's judgment is in one way *secondary*: whereas rehearsing the reasons for some moral claim, in the absence of a genuine spontaneous judgment, does not

---

<sup>89</sup> Compare Driver's discussion of Satan as "an expert in [moral] judgment" (2006, p.630) but one that turns out to be, for rather different reasons ("Father of Lies," and so on), rather unreliable.

amount to moral understanding at *all*; on the other hand, at least some cases of spontaneous moral judgments *do* count as displays of moral understanding, even if one can *not* rehearse the reasons for the judgment; though it may be that a *general* inability to rehearse the reasons *would* threaten the status of those judgments.

**10.** What then of moral testimony and moral knowledge? I have urged—with everyone else—that *something* is not passed on in moral testimony, some ‘moral understanding’; and I have tried to get a little clearer on what it is. My suggestion, then, is that the *uneasiness* that Pessimists feel about moral testimony derives from this. But can moral *knowledge* be passed on by testimony?

It may be that all that remains now is a terminological struggle. Nevertheless, I should like to close by offering a sketch of a view that a Pessimist might take.

Thus in adopting the kind of picture I have, I appear to be implicitly asserting something explicitly denied by, e.g., Hopkins and Driver: both insist that moral knowledge is *propositional* knowledge, and not ‘knowledge-how’. Hopkins’s argument, against moral ‘know-how’, is this<sup>90</sup>: in the first place it is quite clear that, if we reject non-cognitivism, then we must concede that there are moral beliefs, the *truth* or *falsity* of which we care about: moral *propositions*. Now, even if one conceded that ‘knowledge-how’ can be somehow characterized as belief-like, such beliefs would in any case not be ‘beliefs’ that are measured against a standard of *truth*. It is therefore implausible that moral knowledge is knowledge-how.

Now if my own suggestions are on the right track, then there is something, a kind of ‘moral understanding’, which is in effect a kind of ‘know-how’: non-propositional knowledge of

---

<sup>90</sup> See pp. 618-20.

how to use certain concepts.<sup>91</sup> But to say this is not yet to enter the target area of Hopkins's argument. However: while I myself have tried to avoid the question of whether testimony can transmit moral knowledge, nevertheless if some Pessimist, with a view like mine, wanted to assert this, presumably what she would say is (something like) this:

Nothing that we acquire from testimony could be *counted* as moral knowledge if it was not accompanied by the kind of know-how implied in being able to use certain concepts; and thus though there *is* moral knowledge which is propositional ("Their behavior was sexist"), one condition on (counting as) *having* such knowledge is prior possession of a kind of know-how.

The possibility of such a view shows that Hopkins's argument, as it stands, doesn't work: he just does not consider the possibility that there might be some *condition on having* some propositional knowledge, which condition is itself not communicable by way of testimony (whether that further condition consist in 'know-how' or something else). And this opens the way for a Pessimist to assert that, after all, what's wrong with moral testimony is its inability to transmit moral *knowledge*, and not anything about the 'usability' of that knowledge.<sup>92</sup>

---

<sup>91</sup> Hills's view is similarly sympathetic to 'ethical know-how'; but she declines to "press" the point: see p. 105. In likening possession of moral concepts to 'know-how' I am drawing on the (putative) fact that they can't be given necessary and sufficient conditions for their application; though cf. fn. 45.

<sup>92</sup> Driver's arguments against 'moral know-how' has, I think, parallel problems.

### 3 JUSTIFICATION IN CONTEXT

1. My topic in the following pages is what would naturally be called ‘moral justification’. That term, however, is ambiguous in a number of ways, and so some qualifications are in order.

The kind of ‘moral justification’ I am interested in is the kind that would be given in response to (what we could call) the ‘How-could-you-do-that?’ ‘Why?’. Here, some action is queried in such a way that, if no compelling response can be given, it will be understood that the agent has done something *morally wrong*—has offended against the moral law, or the virtues, or some such thing. If some appropriate response *is* given, then we shall say that the agent has ‘justified himself’ or ‘was morally justified’ in what he did.

Now it is a familiar fact from epistemology that, at least in one sense of the word ‘justified’, one could be justified in believing *p*, and yet *p* be false. This sense of ‘to be justified’ is sometimes glossed as, roughly, ‘to have jumped through the necessary hoops; that is, done all that is reasonably required to find out whether *p*’. ‘Justification’ in its moral sense, however, is most often used to mean not just ‘to have jumped through the right hoops’, but also ‘to have done the right thing’. This departure is interesting. Is there an analogue, in the moral realm, to the epistemological ‘justified, but wrong’? Here I think we have to admit that people talk in rather different ways about such scenarios, and it’s not clear just what to say. Perhaps the closest analogue to the epistemological ‘justified but wrong’ is ‘exculpated’ (“It would have been better

for you to have done A, but under the circumstances we can quite forgive you for having done B.”) I don’t want to make heavy weather of this last point; but I bring up the issue because the lack of sensitivity to this distinction in the use of ‘morally justified’ somewhat handicaps me in the point I should like to make. In particular: I should like to describe a rough schema for a kind of *pragmatic picture* of justification without begging any questions in the debate between realists and a certain kind of non-realist in ethics. This difficult task would be made much easier if we could talk about the social practice of ethical justification while remaining agnostic on the question of whether or not there is anything more to being *truly justified*—or as I should like to say, ultimately *right* in one’s conduct. Therefore in what follows I shall use the word ‘justification’ (and its cognates) in a somewhat artificial sense, as a direct analogue to the narrow sense of epistemological justification. Hopefully this will not be seen as in any way question-begging. The problem, very roughly, will be that talking about justification in its ‘realist’ sense—in the sense of ‘actually *right*, regardless of what anyone thinks’—is, I think, just what misleads us when we investigate this concept. For—to give a little preview—using the word in this sense, in the context of *this* investigation, induces us to think that there can only be a finite list of ‘reasons why some action is wrong’, and this is precisely what I wish to deny. Nevertheless, so far from desiring to disparage the realist, I hope to show, when all is said and done, that none of my analysis tells in the slightest against moral realism. However this issue must be postponed for the moment, for it can only be seen in the correct light after further work has been done.

With these distinctions in hand I can now tell my story. In these pages I should like to defend the following thesis: there is an ineliminable, pragmatic, contextual element to the practice of justification. This “pragmatic, contextual” element appears in the following way:

there is no single ‘reason’—no single belief or desire—or group of reasons that is *my reason* or *my set of reasons* why I did any action A; and thus what I *say* in my defense will depend on certain facts about my interrogator, in particular, what he knows and doesn’t know, his expectations, and so on.<sup>93</sup>

It may be helpful, in motivating this thesis, to give a kind of broad sketch of an alternative to the traditional understanding, in philosophy, of the concept of justification. Thus, traditionally, justification has been understood as a kind of abstract, timeless relation between propositions; one proposition is said to *inherit* or *receive* its justification from another, or from the position it holds within a ‘web’ of propositions; justification is said to be *transmitted* from one proposition (or belief) to another; etc.<sup>94</sup> In epistemological inquiries, this model has tended to raise questions about the ‘ultimate source’ of justification, of regresses of justification, and so on.

All of this is perhaps a bit surprising if we look at places where we actually say of someone that he *is justified* or *has justified himself*. In any case, the grammar of the word ‘justify’ is understood, on the tradition, as roughly analogous to the grammar of the word ‘maintain’: transitive verbs that represent ongoing (non-terminating) processes.<sup>95</sup>

My own view is that ‘justification’—as it is ordinarily used—operates much more closely analogously to words like ‘heal’, ‘fix’, and ‘vindicate’. Such words also represent processes—actions, really—which take time; but they differ from (e.g.) ‘maintain’ in several ways. In the first place, they are ‘terminating’ actions: at some point they can be said to have been completed.

---

<sup>93</sup> Although I would distinguish it from these others, my view shares some elements with the ones forwarded by Wellman 1971; Timmons 1999 (see Chapter Five); and the general theory of explanation given in van Fraassen 1980 (see Chapter Five).

<sup>94</sup> Timmons derides this as the “epistemic charge” view of justification (*ibid.*, p.188).

<sup>95</sup> On the other hand, *maintaining* is still an *action*, that is, something done by *agents*; whereas the tradition speaks freely of propositions and beliefs ‘justifying’ things (generally: other propositions). Perhaps then a better model might be the word ‘support’, which indeed is often used interchangeably with ‘justify’ in the literature.

Moreover, although they can be put into a predicate position, with their ostensible subjects dropped (“The bike *is fixed*,” “The patient *is cured*”), they do not seem to operate like such simple property predicates as ‘is red’. In particular, they describe something like a return to a *default* position—though that default position is *not* namable by the predicate itself. Thus two bikes might be in every other way identical, though of one of them we could say that it ‘is fixed’, and not of the other. As we could say: it is not *possible* to ‘fix’ some bikes, and for all that there might be nothing wrong with them.<sup>96</sup>

I am moved by the fact that ‘justify’ operates in many ways like these words—and in particular like the word ‘vindicate’, which it appears to closely resemble semantically as well. In at least one of its uses, ‘justify’ refers to an *action*, which *takes time*, and *can be completed*; when it appears in the predicate position, it seems to be a passive voice construction out of the ordinary, active voice, and not a simple property like ‘is red’; and in the present passive, it cannot be joined with a ‘subject’ term by a prepositional phrase, without generating nonsense (“The claim is justified by Jones”). In all these ways it resembles words like ‘fix’ and ‘heal’ and ‘vindicate’. And if it really is, like them, a kind of *contrastive-status* term, then it will not be possible to ‘justify’ some claims, though for all that there is nothing wrong with them; on the contrary, it will only be possible to justify claims (or perhaps better: justify *one’s claiming something*) if that claim has first been *indicted* or *impugned*.<sup>97</sup>

Of course the foregoing remarks, inasmuch as they have any merit at all, are something like two-edged swords since, it will be said, insofar as it can be made out that the ‘ordinary’ use

---

<sup>96</sup> For this reason it might be better to say that ‘fixed’, ‘cured’—and ‘justified’—name *properties* of objects only in the thinnest sense of the term ‘property’—the same sense that would allow us to say that ‘won’ is a property of some games.

<sup>97</sup> The last paragraphs are a summary of a longer discussion I have tried to carry out elsewhere: “Personal Justification,” unheralded MS.

of ‘justify’ operates as I have claimed, the more it will simply be insisted that, in philosophy, we are simply after something else—a timeless, or anyway ongoing, abstract relation between propositions, or claims, or believables, or whatever. However that may be, my hope is that the foregoing, and what follows, will be something like mutually reinforcing.

2. Let us take up a simple example of a case where I might be asked to ‘justify myself’ in the relevant sense: if I fail, I shall be understood to have done something *immoral* or *wrong*. Suppose I steal some rolls. Some moralist takes me to task for this, and now I have to defend myself. Let us further suppose, for the moment, that it is widely accepted that there are conditions under which stealing rolls is not wrong. The most natural thing for me to say, supposing that it is true, is:

(a) “My family is starving.”

But suppose on the other hand that my interrogator *knows* that my family is starving; to *him* I therefore might have to say,

(b) “I have no way of earning the money (in time) to buy rolls.”

I think that in ordinary life, either of these two claims—depending on the context—would naturally be called ‘my justification’. What does this show? It will be natural to object that:

Of course, when one is pressed for justification, he may *say* only some one thing or another. But if a man is in fact justified, then there must be a finite number of considerations which, taken together, justify him. You have, in fact, produced two considerations which, for the case in question, must both be true for the man to be justified. But this does *not* show, what you need, that there is not a finite list of considerations which, as I say, together justify you.

I have to admit that this line of thought is very seductive. But consider, for example, the following additional considerations that might be adduced:

(c) “(I thought) these rolls were edible”

- (d) “(I thought) rolls would provide sufficient nourishment to prevent starvation”
- (e) “(I thought) no one else would suffer much if I stole the rolls”
- (f) “(I thought) taking these rolls wouldn’t cause animals to suffer”

....

Now it is easy to suppose, as I will ask the reader to, that if any of these considerations had been false—if I had believed the opposite of any of them—I wouldn’t have taken the rolls. Does that show that any of these could have served as my justification?

In the first place, these four considerations can be classed into two groups: the first two have, *prima facie*, no moral content, whereas the latter two are quite clearly morally valenced.<sup>98</sup> But it seems to me that *any* of these four could play the role of justifying me<sup>99</sup>; whether or not they do will depend on the interests and knowledge of my interrogator. Thus if Jones fails for some reason or another to understand the connection between, say, rolls and starvation, I may well have to say to him, e.g., either (c) or (d). And certainly he will not forgive my theft until he understands the connection between it and my goal of preventing starvation. Of course this example is bizarre in the extreme, since just about everyone knows, e.g., that rolls are nourishing and edible. But to make this protest is in fact to grant the central point I wish to make: that what we give in justification for our actions depends on what people know and don’t know. In fact, with a little imagination, we can dream up scenarios in which the condition of my family and my penury are known by everyone, and thus that (c) and (d) would come up more often in my justifications than (a) or (b).<sup>100</sup> And it should be quite obvious that (c) and (d) do not exhaust the potential reasons of this kind. In fact, (c) and (d) arguably demonstrate the way in which the pragmatic dimension of action *explanation* carries over to *justification*, for if Jones had no moral

---

<sup>98</sup> Cf. Foot’s distinction between ‘direct’ and ‘indirect’ moral relevance in “Are Moral Considerations Overriding?”, in her 1978.

<sup>99</sup> Recall that I am using ‘justify’ in a narrow sense.

<sup>100</sup> For example, in a country where everyone is starving, rolls are usually inedible (gone stale), and rolls are often used as door-stops.

qualms about my taking rolls out of a store without paying, but the same general alimentary knowledge, he would most certainly still be asking, “Why did you take the rolls?”, though this time with no rebuke, and the right answer could easily be (c) or (d) or both.

I ought to stress here that the point is entirely general, and does not depend on any quirk in the example. Indeed, items like (c) or (d) could presumably be multiplied indefinitely: in a world where rolls often blew up, or contained arsenic; or where getting from one place to another was very difficult; or where the police state ensured that I should almost never get away with it; in all such worlds I should have to explain to an interlocutor that (as it seemed to me) these conditions did *not* presently hold, and thereby escape his indictment.<sup>101</sup> It would be no good to protest that such contingencies do *not* obtain here; that is in effect my point: what counts as a justification is relative to the expectations of my interlocutor, what is *missing* for him; and that in turn will be in part be a function of the general pattern of contingencies in his world. After all, it is possible, too, to imagine a world in which food was super-abundant, and no one ever starved; a work of philosophy composed in such a world would, I suppose, encounter the same objection over its inclusion of (a) as a possible justification. But that this is no objection is shown by the fact that our world is possible—indeed, terrifyingly real.

What is happening that makes the generation of considerations like (c) and (d) possible? Really only this: there are an indefinite number of contingencies we can imagine which would make an action fail to *come off*; or else make it *pointless*. If some one of those contingencies happens frequently enough, then the justification of a course of action—in the sense of: discharging an indictment—will have to make reference to one’s belief that, this time, this contingency will *not* come up. And really the point is only dramatized by reference to frequency

---

<sup>101</sup> –Supposing, as I think we ought to, that if I (e.g.) *knew* that I was going to be caught by the police, my action would not be justified after all.

of some contingency in a context: this helps us make sense of *why* some interlocutor might be puzzled. But even in *our* world, a man with very strange ideas might ask after (say) (c) or (d).

The considerations labeled (e) and (f) are a little different. The point of including them on the list is to remind us that the morality of some action or other could be queried on any number of fronts. Thus if it were meat, and not rolls, that were lifted, (e) or (f) might be the appropriate response to someone who, e.g., was cool with stealing but had qualms about certain aspects of the poultry industry. It might be argued that moral *progress* is sometimes made by taking actions never before thought to have moral ‘weight’, and asking, in the light of some consideration, whether they’re permissible after all.

Now there may be a temptation here to say that, in cases where (e) or (f) would be the appropriate reply, what is being queried is actually a *different action* from the one we started with—that is, not *theft*, but (say) complicity in inhumanity. In fact, I have no problem with this suggestion, but it seems to me a matter of indifference since, in the first place, my point regarding (c) and (d) would still stand and, in the second, this device, if adopted, would merely have the effect of multiplying *actions* (or action-descriptions) rather than *justifications*, which, from my point of view, comes to much the same thing. It may be thought, nevertheless, that this device has implications for the particularism/generalism debate, though I think this is wrong. It is that debate, in any event, to which we shall now turn.

**3.** I should like to examine an argument made by Sean McKeever and Michael Ridge in defense of a particular species of ‘generalism’. Specifically, I should like to defend the following very weak thesis:

The very nature of moral thought does *not* make it necessary that codification of moral principles be possible.<sup>102</sup>

This is, as I say, a logically weak thesis; and its defense against *one* argument might hardly be supposed to vindicate a particularism of any kind. But I think reflection on McKeever and Ridge's argument and—as I shall argue—its failure, illuminates the particularist/generalist debate quite well, and moreover bears interestingly on the some broader themes. The argument in question can be paraphrased roughly thus:

While there may be someone who is entirely morally virtuous, but unable to articulate the reasons for his choices, the truly *practically wise agent* is someone who always successfully identifies all candidate reasons (for doing or refraining from doing something); successfully identifies the relevant 'defeaters' and 'enablers'—i.e., those considerations which either make a putative reason irrelevant, or a putatively irrelevant consideration relevant; and successfully identifies what there is most reason to do when the candidate reasons are, as a group, equivocal. Now if there could be such a man—and surely we all agree to this—then it must be possible to codify all of morality into principles in the form of conditionals, where the antecedents contain only descriptive terms, and the consequents contain moral verdicts on action. For, whatever the *phronimos* identifies in the way of reasons (and defeaters and enablers), he must identify *all* of them, as we've said, and, after all, no one, not even our imaginary *phronimos*, could identify an infinite number of things, since he is merely a human being. So in fact the number of reasons (etc..) must be finite and, moreover, manageably small in number. Therefore these 'reasons' could be put into the antecedents of conditionals, and we would end up with a system of manageable moral principles.<sup>103</sup>

Now the argument as it stands is not complete, since, for example, it is not clear from the foregoing why we should believe that the antecedents of these conditionals could really be reduced to 'purely descriptive' terms. And if the antecedents of these principles were to contain 'thick moral terms', the generalist will have shown very little: if we assume the unity of the

---

<sup>102</sup> This is, effectively, what McKeever & Ridge label "Anti-Transcendental Particularism," in their 2006; see pp. 15-16.

<sup>103</sup> Adapted from McKeever & Ridge, pp. 140-144.

virtues, as has traditionally been done by virtue ethicists, then, e.g., “If an act would be kind, do it,” is fairly uncontroversial; obviously the difficulty lies in deciding which cases are cases of kindness, and the particularist will merely assert that the difficulty, just swept under the rug, re-emerges here. This point, which scarcely needs saying in the present context, is perhaps rather important in another; but we shall revisit it in the end.<sup>104</sup> Suffice it to say, McKeever and Ridge advance a different argument, a little later, to patch up this hole—or rather, a reiterated variation on the same argument. We shall return to this issue as well.

4. One is impressed with the feeling, rehearsing this argument, that he has witnessed some sleight of hand; and I think this impression is justified. But what has gone wrong here?

Well, what are we being asked to imagine when we are asked to imagine someone who always “successfully identif[ies] candidate reasons” and “successfully identif[ies] what there is most reason to do”?<sup>105</sup> This is as much to say, “*God* always knows what the right thing to do is—and the right reasons for so doing. What those reasons are we can’t say, but that is another matter.” But even *God* would have to convince us.

I can’t speak for the reader, but when I try to imagine the *phronimos* as here described, I imagine a man standing in front of a painting, a depiction of a ‘moral situation’, and the man draws circles around various parts of the picture—the relevant candidate reasons. But of course this won’t quite do, since, e.g., ‘the fact that she wasn’t coerced’ might be a relevant reason, and quite clearly this can’t be circled in a painting.

---

<sup>104</sup> It is tempting to say that this point—that morality can surely be captured in principles spelled out in thick terms—would be denied by no one; but actually Dancy, the arch-particularist, does deny it (see his 2004, p. 84-5; 118-123). For a defense of a kind of soft particularism, which allows that morality can be codified in rules that make use of thick moral terms, see McNaughton and Rawlings 2000. I am largely in agreement with the conclusions of their paper, though I am less sure about the arguments.

<sup>105</sup> *Ibid.*, p. 140.

Now the fact that *lacks* can be reasons is not strictly necessary for our point, but it illustrates the point rather dramatically. That point is, in fact, just the one made above (§2), namely, what counts as a reason for or against some action is, in part, a function of the interests, knowledge, and demands of one's interlocutor; and the limits on these latter are amorphous, shifting, and, most important, not delimitable *a priori*. If this is right, then clearly morality is *not* necessarily codifiable.<sup>106</sup> But it is worth working through some of the details to see why this is so.

We can usefully begin with the example from §2 above. How might we formulate a principle indicating when it is permissible to steal? Supposing, as we did above, that there is at least one kind of exception to the general prohibition, and passing over the problem of how the ostensibly 'thick' term 'steal' might be cashed out in descriptive terms, we might begin with:

(P<sub>1</sub>) It is wrong to steal, unless one is starving and needs food. (I.e., if an action would be an act of theft, don't do it, unless one is starving and needs food.)

This is a pretty good principle, and I think I'd more or less endorse it for use in daily life. However, it will not suit the generalist's needs as it stands. In the first place we had better amend it to:

(P<sub>2</sub>) It is wrong to steal, unless one is starving and needs food, in which case one may steal food.

This qualification having been made, we may recognize that P<sub>2</sub> is still ambiguous, and thus feel compelled to rewrite it as:

(P<sub>3</sub>) It is wrong to steal, unless one is starving and needs food, in which case one may steal food, provided that that food is edible and capable of nourishing.

---

<sup>106</sup> Which is not at all to say that morality is necessarily uncodifiable. It is hard to put this point in a non-tendentious way, but I should like to say: if men in the future should decide that utilitarianism was the correct moral view (and *not*: the correct moral theory), then they should have to say that morality was codifiable. Perhaps it would be best to say: the truth of (principled) particularism or generalism could only follow from our moral views, and not precede them.

It needs no mentioning that this principle is rather ridiculous; and yet P<sub>2</sub> really was in one way incomplete: if I was right in what I argued above, then considerations (c) and (d) can sometimes be reasons which justify my action. And if they are then they will need to be referenced in any ‘complete’ principle. For clearly the fact that I am starving does not license me to steal food which I know I cannot eat. In which case references to conditions (e) and (f) will have to be made as well. Thus we shall have:

P<sub>4</sub>) It is wrong to steal, unless one is starving and needs food, in which case one may steal food, provided that that food is edible and capable of nourishing, and provided that the theft does not result in significant suffering for anyone else, nor in the suffering of animals.

And so on. Conditions (e) and (f) will have to be built into any principle concerning theft to protect it from what we might call ‘novel moral objections’. It might be argued here that we do not need to add riders for (e) and (f), but only add something to the effect of, “and so long as, in doing so, one does not violate any other moral rule.” Perhaps that can be made to work; but it expresses an optimism about *the number of other rules* that might seem, in light of the present discussion, misplaced.<sup>107</sup> I will not pursue that question here. In any case it is irrelevant since, if the discussion in §2 was sound, conditions like (c) and (d)—rather esoteric, apparently ‘non-moral’ conditions—can be multiplied indefinitely.

**5.** McKeever and Ridge disparage the possibility that esoteric facts—such as: the color of one’s shoelaces—could be morally significant. In fact they have an argument to the effect that such considerations cannot be relevant, and I will consider it shortly, although I must admit that I am not sure I understand it. The question to which we need an answer is: what makes a

---

<sup>107</sup> It must be remembered that these other rules are all, also, to be spelled out without the help of any ‘thick’ moral terms.

consideration ethically significant? I have just argued, effectively, that nearly any consideration can be ethically significant in the right context. On the other hand, it must be admitted that some considerations seem to have a *prima facie* ethical significance which other considerations lack. McKeever and Ridge consider the possibility of a distinction between “primary” and “secondary” reasons,<sup>108</sup> before ultimately rejecting it. But others have found some usefulness in similar distinctions, e.g., between “default” and “non-default” reasons,<sup>109</sup> or “direct” and “indirect” moral significance.<sup>110</sup> One tempting possibility which is suggested by the general tactics of this essay is to align the distinction between secondary and primary reasons (as I shall call them) with the distinction between those considerations which could be morally relevant in certain contexts, but generally are not, and those considerations which are almost always morally relevant.

This characterization is perhaps rather thin; but as I shall put no weight on it in what follows, it hardly matters. The point I do wish to stress in the present context is that, if a distinction of this sort *is* tenable, it would nevertheless be a mistake to think that primary reasons would remain primary reasons in all worlds. For what counts as a primary reason may be relative to a culture—a fact which, I hasten to add, gives no aid or comfort to any ‘ethical relativism’. Consider the following case:

Suppose I am being pressured by the high school cross-country coach to join the team; but I hate racing, particularly because I hate the anxiety induced by these kinds of competitive events. Incidentally, one of my moral failings is an over-strong desire to have other people like me. Finally one day I do join the team, run the race. And my friend J says, “How could you do

---

<sup>108</sup> The distinction comes from J.L. Urmson’s “A Defence of Intuitionism” (1974) cited in McKeever and Ridge 2006 at 130.

<sup>109</sup> Dancy, e.g., in his 2004.

<sup>110</sup> Philippa Foot, *ibid.*: see fn. 98.

that? How could you be buffaloes into running for the team? That’s so cowardly of you.” And I offer in justification: “You know what? It suddenly struck me, I really did want to race. I just felt a powerful desire to compete and win.” A scene like this is admittedly not terribly common. (Although they do happen.) In fact, the rarity of such scenes is important. But the present point is that it is easily imaginable, it could happen. Now imagine a tribe—“high school students”—for whom *not being buffaloes into stuff* is very important and in fact comes up all the time. If this tribe should develop its own moral vocabulary, the following might be one of their principles: don’t do anything that you don’t want to do. It’s worth noting that this principle is defeasible. But we mustn’t imagine that there are more exceptions to this rule than instances of it; again, we must imagine that “getting buffaloes” is something that comes up all the time in this culture, and therefore something to guard against.

Now in the first place we should note that we—we non-high school students—can accept and agree with the “moral outlook” of this tribe: after all, it is cowardly to be steamrolled into things one doesn’t want to do. But their principle will be no principle of ours, since, in our own lives, it is probably (much) more often important to do what one does not want to do, than to refrain from doing what one doesn’t want to do. What does this show?

I think this shows that principles are relative to one’s circumstances—or rather, are relative to the larger patterns and likely contingencies of one’s life. So the idea of “the principles that the truly practically wise agent follows” or “the principles that we can read off from the behavior of the practically wise agent” makes no sense.<sup>111</sup> This is the problem with merely supposing an imaginary practically wise agent who “always successfully identif[ies] candidate reasons”: ‘reasons’ are not helpfully thought of as features of a situation, but rather as *things we*

---

<sup>111</sup> I do *not* mean to imply that no one ever thinks in terms of moral principles, or even that it is a bad thing to do so.

say to each other to justify ourselves. And what we say to someone to justify ourselves will depend on all sorts of facts about *that other person*—in particular: what is salient to her, what sort of contingencies she's on the lookout for, what she takes for granted and what she doesn't, what the major variables are for her in ethical situations. And this is what I meant by "Even God would have to convince us."

6. Now McKeever and Ridge consider, what they acknowledge would be fatal to their program, the possibility just defended, that "any consideration can be a reason." And they reject this possibility on the basis of a distinction between 'a reason an action is wrong (right)' and 'a reason to believe an action is wrong (right)'. Like most distinctions, this one could clearly be put to some use or other. But in the context of McKeever and Ridge's discussion, it looks like straight-forward question-begging. Thus, they write:

Jonathan Dancy suggests that in the right contexts the fact that someone asks you for the time is a reason to tell her. In our view, the fact that the person asked you for the time is a reason to believe you ought to tell her and not itself a reason to tell her. The reason to tell the person presumably is something like the fact that telling her would be polite, or the fact that it would satisfy some desire of hers, would undermine her anxiety, or would help her achieve some (morally permissible) end. If telling the person the time would do none of these things then there may be no reason to tell her even though there may still be reason to believe that you ought to tell her. Of course, we could count the fact that she asked as a further reason to tell her in addition to all of these other reasons, but this seems like one reason too many....<sup>112</sup>

In the first place, this example is actually quite useful, since it minimizes the temptation to imagine a deliberative process preceding one's action—no one thinks *at all* before giving the time—and therefore minimizes the temptation to suppose that one's 'reason for action' was whatever one considered just before acting. But I must confess that I don't entirely understand

---

<sup>112</sup> McKeever and Ridge 2006, p. 126.

the rationale for the move being employed here. On my view, of course, *all* of the listed potential reasons could be a reason to tell Robinson (as we might call her) the time—what one would actually *give* as his reason would depend on the context. But I cannot understand how one could deny that Robinson’s asking me the time is a reason to tell her. Certainly, it would make a fine answer to the question, “Why did you just tell that woman what time it was?” I think it is possible that what is moving McKeever and Ridge here is the fact—which I gladly concede—that the situations in which one would explain his having told someone the time by saying, “She asked me the time,” are few and far between. But, as I have said, this is just the point I should like to make.

Now there is something misleading about the list of potential reasons given in the quoted argument, since that list contains “the fact that it would satisfy some desire” of Robinson’s; and then McKeever and Ridge go on to say that “if telling the person the time would do none of these things then there may be no reason to tell her.” Surely if Robinson has no desire to know the time, but asked me anyway (why?), then there may be no ‘objective’ reason to tell her: it would be pointless. But presumably this is a red herring.

In fact, McKeever and Ridge reject the possibility that one’s “conscious thoughts”<sup>113</sup> are dispositive of one’s reasons for acting, as indeed they must, since one may think of all sorts of things just before one acts. They note that if (enlivening her character) Robinson has promised to buy her friend a plaid cap, she may only think—i.e., have the “conscious thought”—before buying one, “This one’s plaid, so I’ll buy it.” But then they write:

Her conscious thoughts might instead be understood as enthymatic [*sic*]. When fully spelled out this argument might instead go something like, “This baseball cap is plaid, I promised to buy my friend a plaid baseball cap, I have good reason to fulfill this promise, so I ought to buy it;

---

<sup>113</sup> *Ibid.*, p. 130.

therefore I will buy it.” One reason for understanding [Robinson’s] thoughts in this way would be that she would articulate the premisses of the argument if pressed for an explicit justification.

But of course, this isn’t actually true. What Robinson would say would depend on what she was asked. And in some situations there would be many other things she would have to add to her “explicit justification”—for example, “And I have enough money now to feed my starving children,” and perhaps, “I share the execrable sartorial sensibility of my friend.”

I think McKeever and Ridge actually give the game away in the next paragraph, where they write:

Eccentric facts like the fact that one would be buying a plaid baseball cap for someone are hardly guaranteed to make the action seem morally attractive to a virtuous agent. Imagine trying to explain to a morally wise but puzzled agent *who does not know about your promise* why you are buying a plaid baseball cap simply by saying, “Look, this is the buying of a plaid baseball cap! Oh, and no other feature of the situation explains why that fact is not a good moral reason to buy it.”<sup>114</sup>

Here McKeever and Ridge are making use of exactly the device which is the lynchpin of my view, namely that what one gives as a reason depends on what one’s interlocutor knows. Of course, McKeever and Ridge are right that Robinson’s justification (or explanation) of her action here is terrible; but this is precisely because her wise but bemused interlocutor is only told what she does know, rather than what she doesn’t.<sup>115</sup>

In fact this example is not quite right for the generalist’s argument, since, as it is given, it is not clear what Robinson is supposed to be *justifying* herself about—except perhaps her bad taste. Moreover, the discussion was supposed to be about the plausibility of “This is a plaid baseball cap” being a reason for action, not “This is the buying of a plaid baseball cap.” Emending the example then, we would have:

---

<sup>114</sup> *Ibid*, emphasis added.

<sup>115</sup> Their use of the word ‘guaranteed’ is also illicit, though perhaps telling.

Imagine trying to justify yourself before a morally wise but puzzled agent—who 1) knows about your promise but 2) thinks ‘plaid’ means tattersall and *vice versa*, and 3) is now taking you to task for your purchase—simply by saying, “Look, this is a plaid baseball cap! Oh, and no other feature of the situation explains why that fact is not a good moral reason to buy it.”

Guarantees are of course neither here nor there; but this looks like a pretty good justification.

I think this gives us good ground to reject—in this context—the distinction between ‘a reason an action is wrong (right)’ and ‘a reason for believing an action is wrong (right).’ One is tempted to suppose that McKeever and Ridge employ this distinction merely because it suits their purposes; at at least one point, their language seems to betray this.<sup>116</sup> But again I acknowledge that I may be missing something here.

7. It should now be clear why I have employed this (somewhat) artificial sense of ‘justification’ in the foregoing discussion: to speak, as the alternative would demand, of ‘right-making features’, would only encourage the intuition that, whatever one may *say* in his defense, his action will only be right if it is *made* right by some ‘feature(s)’ of the situation; which, in turn, encourages the intuition that the principles of morality must be codifiable. Thus, McKeever and Ridge, considering wrong-making features, write, “[N]obody seriously contends that our senses tell us everything. So the features of a situation in virtue of which we just see that a given action is wrong must themselves be limited.”<sup>117</sup> Strictly speaking, the quoted argument is invalid, relying as it does on the illicit inference from “It is impossible that we see everything” to “What we see is finite.” I suspect that the apparent plausibility of their conclusion trades on a particular—a problematic—conception of “features of a situation.” In fact, I have no objection to talk about ‘right-making features’, ‘wrong-making features’, or ‘features of a situation’ quite

---

<sup>116</sup> *Ibid.*, p. 129: “Even if the thought, ‘this cap is plaid’ did figure in the deliberation...we need not understand it as her reason for action. We could instead understand that thought as her reason for belief.”

<sup>117</sup> *Ibid.*, p. 115.

generally, so long as it is borne in mind that the ‘features’ of a ‘situation’ are infinite in number, being, as they are, merely reflections of what one might possibly say in successfully defending himself.

As the last sentence indicates, my picture involves a kind of reversal, of one aspect anyway, of the picture employed by many moral philosophers. For whereas McKeever and Ridge evidently imagine a world of ‘features’, some of which are morally relevant, and a *phronimos* who reads these latter features off the world and then acts on them,<sup>118</sup> I should prefer to say that what counts as a morally relevant feature of a situation is read *back into* the world after it appears as the content of a justification.

**8.** Where does this leave our thought experiment? The challenge posed by McKeever and Ridge was: the bare possibility of a truly practically wise agent entails the codifiability of ethics. But I have argued that morality is *not* necessarily codifiable.<sup>119</sup> Must I now deny that the truly practically wise agent is possible? Before answering this question we must look once more at the thought experiment, in the new light of the foregoing discussion.

McKeever and Ridge imagine the *phronimos* as “successfully identifying the candidate reasons” and “successfully identifying what there is most reason to do.” Now ‘successfully identify’, like ‘know’ is factive; and therefore we are tempted to think that, whatever a man may say in justification of himself, his action is *made right* by some ‘feature’ or other of the situation; and since a man, as the limited creature he is, can only ‘identify’ so many things, it looks as though, if a man—the *phronimos*—can always ‘successfully identify’ the features that make an

---

<sup>118</sup> “[R]easons themselves are simply descriptive facts which favor not performing the action,” p. 115.

<sup>119</sup> —While conceding that, for example, if we all agreed that utilitarianism was correct, generalism would follow trivially.

action right (or wrong), these features can be listed as the antecedents of conditionals which just are the true moral principles. But if my argument up till now has been sound, then to imagine the *phronimos* is to imagine a man who, should he say of some action, “This was the right thing to do,” will, no matter what we should reproach him with, always *come out on the right side*. That is, he will be able to convince us, for whatever consideration we should raise, that that consideration does not count decisively against the action. And I should like to say: *this is quite imaginable*.

It will help here to put the argument in its most direct form. Thus the generalist will want to say:

We can willingly grant—though one must be rather cynical to say so—that some sophisticate may be capable of always being ‘justified’—in *your* sense of that term—in ethical matters. But we were not talking about *this* kind of justification. We were supposing that the *phronimos* is always *truly* justified—justified *regardless* of whether or not anyone *thinks* that he is. After all, an action is either right or wrong; and *these* facts transcend what anyone has to *say* about the matter. So all your talk about the relativity of ‘justification’ is simply irrelevant here. For if you admit that facts about right and wrong transcend what anyone thinks—as you must, on pain of endorsing one of the most callow and jejune of anti-realisms—you must further admit that what makes some action right or wrong *runs free* of what anyone says to justify himself—in fact is logically independent of the latter.

Here again we see the difficulties that come from employing the ‘wide’ sense of ‘justified’. What we should all admit, I take it, is that whether someone has *acted well* or *acted badly*—done the right thing or the wrong thing—is independent of what he says to us. Someone can have done the right thing and convince no one that he is not a sinner; or done the wrong thing and convince everyone that he is a saint. If we choose to use ‘justified’ in the wide sense, it will follow that someone can be ‘justified’, too, though he has convinced no one—just supposing he has rehearsed—or: is aware of?—the ‘right reason’. But, in light of the foregoing discussion, the idea that there is some one, or some few, ‘right reasons’, which suffice to justify a man when he

has (e.g.) been indicted for theft; this idea looks more and more like an unsupported hunch. Just what would single it out?

I myself cannot see the point of a concept of moral justification which cuts it off so radically from *what people actually say to each other*. I think we should insist that, when we are asked to imagine a man who is always right in moral matters, while prescind from any reference to the practical activity of justification, we no longer have any idea what we are imagining. *Our judgment of him as being right in moral matters must go through the social practice of justification*. And therefore—what in some ways is more obvious—our moral principles must *reflect* this social practice of justification. And so, if my argument till now has been sound, and this practice is inflected by contextual and pragmatic considerations, our principles must reflect these considerations in turn.

For these reasons, it seems to me more clear to ask, not whether the *phronimos* is possible, but what we should say of someone who appeared to be the *phronimos*. The first possibility, at which I have already hinted, is that we may suppose, with the cynic, that the putative *phronimos*, while considerably good at justifying himself, may yet be *wrong* in at least some of his judgments—even though we can't say why. This is to say that this '*phronimos*' may be no *phronimos* after all, but a sophisticated fraud; and therefore no explanation of his 'ability' may be necessary.

Now to endorse this view of the '*phronimos*' is *not* yet to plump for callow anti-realism. In fact the callow anti-realist, having conflated 'justification' in my narrow sense and 'justification' in its ordinary sense, cannot even take such a view; all she can say is that the '*phronimos*' is considerably good at justifying himself, and *therefore* right in what he does.

What the callow anti-realist shares with the cynic is an explanation for the apparent talent of the *phronimos*: it is a talent for argument.

But there is no reason to be so cynical. One possible explanation for the *phronimos*'s talent might be sketched along the lines adumbrated by Professor McDowell. Thus if McDowell is right, and the rules governing the use of thick moral terms like 'cruel' and 'charitable' can not be cashed out in purely descriptive terms—because there is no pattern at the subvenient level—then it would be possible to reconcile the *phronimos*'s talent—he just sees cases of kindness and cruelty—with my argument against the necessary codifiability of morality.<sup>120</sup> But the *phronimos*, then, will display his talent for justification, not in giving some one knockdown reason; but by being able to answer any possible query put to him.

Against this kind of “thick intuitionism”<sup>121</sup> McKeever and Ridge redeploy a version of their original argument for the codifiability of ethics: any thick moral term must supervene on descriptive facts; if someone has competence with a thick moral term, then he uses it exactly where the right descriptive features are present; and for someone to be so competent, there must be only a manageably small number of descriptive terms which fix the use of each thick term.<sup>122</sup>

It is hard to square this thought with any intuitive picture of the thick moral terms we actually use. But in any case, if the argument against moral principles above was sound, it will work equally well here. In fact, arguments over whether or not some act was permissible (“I

---

<sup>120</sup> See McDowell 1979 and 1981.

<sup>121</sup> The term derives from McNaughton and Rawlings 2000, quoted in McKeever and Ridge at p. 147.

<sup>122</sup> See McKeever and Ridge, p. 147-8.

stole the food, but was that wrong?") can often be recast as arguments over the correct application of thick moral terms ("I took the food, but was that theft?").<sup>123</sup>

This move may well spur the generalist to ask, in parallel fashion, how it could then be possible for someone always to recognize cases of cruelty. I shall not try here to recapitulate the relevant McDowellian arguments here—but only add that the force of this question derives its plausibility, I think, from the idea that there is a kind of metaphysical 'sameness' from which all other 'samenesses' must derive: it is not *prima facie* obvious why it should be more difficult to apply and reapply (say) 'courageous' than (say) 'danger'.

**9. A codicil.** While we are on the topic, we should consider one further generalist objection to the possibility of uncodifiable ethics. Thus it might be said:

You've made reasons relative to context and world. But couldn't one say that principles followed by the *phronimos* would merely vary from context to context, and world to world? That is, why not suppose that that each context and world just has its *own* principles; and that the truly wise practical agent, sensitive to contexts and world, merely changes his principles from context to context accordingly?

Something like this could perhaps be said; and it even seems to capture something of our ordinary experience of moral reasoning. The problem, as I see it, is that a move like this merely shifts the problem, from infinitely long principles to an infinite number of rules. It seems to me a matter of decision whether we say, for example, that we have one rule regarding theft or many—although it is a decision that would, again, have some consequences: as I say, it may be a matter of psychological fact that we often think in terms of rough-and-ready rules, and, for example,

---

<sup>123</sup> This is not to say that this choice will not have consequences: if we think it makes sense, e.g., to admit that some action falls under the domain of a particular virtue (say, benevolence), but then ask whether or not it was really the right thing to do, we may have to deny the unity of the virtues.

teaching our children one hundred rules or ten may make some psychological, and therefore some moral, difference.

## 4 THE METHOD OF MORAL HYPOTHESIS

1. What I want in this chapter to argue is, in effect, that working towards a general moral theory makes about as much sense as working towards a general theory of jokes. Now some people might think a Unified Joke Theory might be a good idea; even I have to concede that it would interest me in a manner of speaking. But there are a lot of reasons to think a General Theory of Jokes would not be a very good idea; that, in any case, there are a lot of things such a theory could *not* do; and, I shall argue, many of these considerations carry over to the idea of a general theory of morality.<sup>124</sup>

Suppose we came up with a Unified Joke Theory; what might it look like? We can imagine such a theory consisting in a system of hedged *principles*: “a joke is funny if it..., so long as it doesn’t..., or ... [and so on].” I can dimly imagine at least some of these principles.<sup>125</sup> How might such principles be devised? Presumably, we look at things we think are funny, and we apply something like Mill’s Method of Differences: why is *this* joke funny when *that* one isn’t? Such a procedure might yield such general principles, as I say; but then of course the principle will always be in danger of being undone by some counter-example (a case, e.g., where

---

<sup>124</sup> Some of what I say below might seem rather glib; but I hope that by the end, I will, not so much have dispelled the feeling of silliness, but transferred it to my target: moral theorizing.

<sup>125</sup> I have heard it said that a joke is funnier, all things being equal, the more *specific* it is. That actually strikes me as a fair general principle.

a non-specific joke is funnier than a specific one). Not that any such example would *ruin* the principle: we might merely have to hedge it further. It will be in our interest then, in devising such a theory, to find as many examples as possible, before pronouncing any principle correct.—And even then, we shall have to allow that the principle might not be correct after all.

On the other hand, there is another method for dealing with counter-examples: if we cannot see how some one *apparent* counter-example is really to be distinguished—cannot, that is, discover the appropriate hedge for our comedic principle—then it would be possible, anyway in theory, to suppose that the error lies *not* with the principle, but with the joke: however much it may make us laugh, the joke is not—it transpires—really funny.

I submit that this would be a ridiculous procedure. Evidently, part of what makes something funny is: that *we find it funny*. On the other hand, this last thought is consistent with the following thought: we can be *wrong* about what's funny. That is one big step towards the notion of *objectivity* in humor. And I think that we should accept that this is not merely a notional possibility, that we do, in fact, believe in *some* objectivity in humor; for we do, at times, retract our claim that some joke was funny. Sometimes, we do this when the joke no longer elicits a laugh. But here it is worth noting that we *don't* always react to such situations merely by saying, "That joke isn't funny anymore." Sometimes we say: "I thought it was funny, but it wasn't, really." And there are other cases, too. Sometimes we *still* laugh, but deny that it's actually funny. Here we say, "I can't help laughing, even though I know it isn't funny."<sup>126 127</sup>

---

<sup>126</sup> It might be objected here that this most often happens when we find something funny which, at the same time, we regard as in some way *morally* offensive; and that therefore this element of the analogy cannot carry over. But this isn't, in fact, the only kind of case: "I never find physical comedy funny, I think it's stupid; I must be laughing because I'm in a giddy mood."

<sup>127</sup> Other considerations also point towards a belief in 'objective humor': when we call someone 'humorless', we do not mean that he doesn't laugh at many things; we mean that he is *missing* something—i.e., wrong.

Nevertheless, there are other uses—uses besides overturning our existing comedic judgments—to which we might put our Unified Theory of Jokes. So, for example, we might wheel in the Unified Theory in order to *convince someone else*. Thus, if Jones doesn't find my joke funny, I might deploy the Unified Theory, and *explain* to him why it is. Now it is a notorious fact that a joke cannot be made funny by being explained; that explaining a joke will usually, in fact, spoil the joke for *everyone*. Does that mean I must be *wrong* about the *reasons why the joke is funny*? Of course not; but it does at least seem to show that the Unified Theory will be useless for this purpose, too.

Here it may be well to remind ourselves that, at least occasionally, we are *unsure* about whether a joke is funny; our judgment on the matter is somehow *suspended*. Could the Unified Theory help us out here, by telling us what to think about novel cases? Suppose the theory tells us that some novel joke *is very funny*; could this decide the matter? I suppose this could have the effect of having us *say* that the joke was funny; but it is hard to see what further consequences it could have. If we don't find ourselves spontaneously *laughing*—I leave aside the interesting and by no means irrelevant phenomenon of *fake laughter*—then, evidently, whatever we might *say*, the Unified Theory has failed to *alter our judgments*. Here, too, then, the Theory is less than obviously useful.

We might contrast this use of the theory with another, similar method, that does not invoke *principles* of humor at all. For we can and do sometimes get a friend to *get a joke*, and this does not always consist in (e.g.) filling in 'background information'. Obviously, there are *many* things we can do here: we can relate the joke to another one; try to emphasize certain aspects of it which may have gone under-appreciated; and even, sometimes, give the *reasons*

*why* it's funny. Any of these tricks might succeed or fail; but of course, failure will not convince us that the joke isn't funny after all. "It's too bad for him he doesn't see it," we might say.

This last point may remind us that, insofar as we *do* believe in something like 'an objective funny', the Unified Theory project might nevertheless have a purpose after all.

Yes [it might be said], the Theory will be powerless to overturn our 'comedic intuitions'; it will in all probability convince no one; and it cannot tell us anything about novel cases; nevertheless [— and here one must switch to the first-person singular—] *I* know when a joke is funny. [Or perhaps: "I and my like-minded friends in my social set."] And it would in any case be an interesting project to discover just *what makes a joke funny*. After all, it is a bit of a mystery; and why *couldn't* there be a set of enumerable principles which give necessary and sufficient conditions for a joke's being funny? In fact, there *must* be such a set, given that we are all, after all, 'finite creatures', and thus we must be operating with *some* set of principles.

The last line contains a mistake, I think.<sup>128</sup> But at the moment I want to raise a question about this semi-aesthetic project. Suppose we grant the possibility that someone's sense of humor could be codified in principles; suppose we even grant that the general sensibility—and hence the principles as well (?)—are widely shared. Still, one might ask, what would the point be, when we know full well that, in twenty or thirty, or anyway one hundred years' time, this sensibility will have been wiped from the face of the earth?

What I emphatically do *not* mean, here, is that humor is, after all, and in the face of *history*, a subjective affair. For it would be possible, e.g., to think that (say) older American cinema evinces an *unsophisticated* sense of humor; or, supposing one is old enough, to regard the contemporary Anglo-American comedic sense as a *degeneration*. (I have heard both claims made.) So there is no barrier to thinking that, if, in a hundred years, everyone is going to have a

---

<sup>128</sup> See Chapter 3.

much different sense of humor, then nevertheless, they will be quite wrong—will have degenerated, or forgotten what is really funny, or something of the kind.

My point is, rather, that, *however* we choose to regard our descendants (or our forebears), our Unified Theory will be *useless* to them, as they will reject it, root and branch. And I think that that must take some of the interest out of such a project. Needless to say, the problem is compounded by the (obvious) fact that *other cultures presently living* will not find anything of merit in our Unified Theory; and that, as it happens, there are probably wildly diverging senses of humor even *within* (e.g.) American culture.

There is of course one way to avoid this difficulty, where the aim is to make the Unified Theory so close to contentless that it will accommodate just about *any* sense of humor. I think there is something to such a suggestion, and I will return to it in the end. For now I only note that such a ‘theory’ would not look very much like, nor satisfy many of the ambitions of, the kind of thing we began with.

2. My aim here, of course, is to try to transfer our skepticism about the Unified Theory of Jokes to the ‘moral theory project’. In particular, I shall be arguing that a moral theory, like our imagined theory of jokes, is a doubtful ground for overturning our spontaneous verdicts; is, at best, just one among many methods of persuading others about moral conclusions, and not a terribly effective method at that; would be a more or less arbitrary way of determining the ‘answers’ to novel moral cases; and, as a mere systematizing project, is likely to be of merely parochial interest. It is to the first of these points that I shall devote the most attention; firstly because I think it is the most important; but second because I believe many of my suggestions here will transfer in an obvious way to the other points.

But before making this case, it will be worthwhile to pause and ask just what a ‘moral theory’ is, or anyway, what I shall mean by it in the following. For many things have gone under this name, and evidently the term is no longer transparent. Thus, e.g., I do *not* mean by ‘moral theory’, “any and all meta-ethical theorizing.” For one thing, I take it that I am currently engaged in *some* kind of meta-ethical thinking myself in these pages; and, more seriously, nothing I say here is meant to cast doubt on the project of getting clear on our *ethical concepts*. That seems to me a real task and perhaps a pressing one.

No, what I mean by ‘moral theory’—or, better, ‘moral theorizing’—is this: the task of trying to bring our various spontaneous moral judgments into some systematic, unified whole, typically, by bringing individual judgments under principles, and then those principles under more general principles still; and in particular the version of this project in which it is accepted that some of our spontaneous moral judgments will have to go, under pressure from the systematizing requirement.<sup>129</sup> We find this project going on, from *one* end, in theorists who begin by deploying some very general formula, a kind of master principle of morality, and try to work their way down to particular judgments<sup>130</sup>; these theorists are generally acknowledged to be engaged in ‘moral theory’. But the term—‘moral theorizing’—could be applied with equal justice to those theorists who work from the other end: who begin with some set of spontaneous moral verdicts (‘moral intuitions’) about ostensibly similar cases, and try to develop some

---

<sup>129</sup> I take this is also what Professor Setiya has in mind in his “Does Moral Theory Corrupt the Youth?” (nd); he defines his target there as “the kind of theory that takes our moral intuitions seriously as starting points and aims to produce a systematic body of principles that vindicates these intuitions by endorsing them, undermines them by failing to do so, and yields justified claims where they [those intuitions] are silent” (pp. 3-4). Cf. also Thomas Hurka, quoted in Setiya (p.4), writing of the project of the early 20<sup>th</sup> Century British moralists: “These theorists shared the general normative project of systematizing common-sense morality, or finding more abstract principles that can unify and explain our particular judgements about right and wrong. [...] only if its judgements could be systematized by a few fundamental principles would they be properly scientific” (Hurka 2004).

<sup>130</sup> Recent examples include Scanlon 1998; or Parfit in his *On What Matters*. (See esp. p. 305 of the latter in the MS for a statement of what moral theory is up to, and its connection with ‘reflective equilibrium’. Both philosophers call their theories ‘contractualism’, and this is indeed one form the big project sometimes takes.)

principle which will divide the cases according to our intuitions. What is again vital, for my point, is the willingness of these theorists to abandon some recalcitrant ‘intuition’ if it fails to fit the general pattern. Of course (as we shall see in a moment), the offending intuition is not merely abandoned; some attempt is always made to ‘account’ for it. But, as I shall argue, there remains something problematic here nonetheless. My aim, in what follows, will be to suggest some grounds for Annette Baier’s suspicion of “the whole idea of a moral ‘theory’ which systematizes and extends a body of moral judgments, and [...] in particular the idea that a theorist might accept a theory with controversial implications, without thereby becoming a moral reformer, one dedicated to having these implications endorsed and acted upon.”<sup>131</sup>

3. In a recent paper<sup>132</sup>, Judith Jarvis Thomson revisits the infamous Trolley Problem, and concludes that, in fact, and in spite of her own previous arguments to the contrary, what “explains the difference between our verdicts” in the various modulations on the Problem is Philippa Foot’s old Doctrine of Doing and Allowing—the distinction, that is, between negative and positive duties, and the corollary that the former duties are much weightier.<sup>133</sup> In the course of arguing for this conclusion, Thomson is compelled to dispense with one apparent counter-example: the ‘common-sense’ verdict in the case of a bystander who must choose either to do nothing and allow five men to die, or to turn the trolley, thereby killing one. (Following tradition, we may call this case ‘Bystander’.) Thomson now argues, following Alexander Friedman, that the common verdict is just wrong: we may not, in fact, take the second option and turn the trolley

---

<sup>131</sup> “Doing Without Moral Theory,” p. 33; in Clarke and Simpson (eds.), *Anti-Theory in Ethics and Moral Conservatism*.

<sup>132</sup> “Turning the Trolley,” *Philosophy and Public Affairs*, Fall 2008.

<sup>133</sup> Foot helped revive the Doctrine in the contemporary casuistical literature, but the general idea—that negative duties are more stringent than positive ones—is of course traditional, and was noticed by nearly everyone. (The name—‘The Doctrine of Doing and Allowing’—originates, I believe, in Quinn 1989.) Although I go on to disparage the use to which Thomson puts the principle, it is clear enough that the core of the idea is sound.

in such cases. Then why, one will ask, does everyone think we may? Here Thomson suggests that our intuitions go awry in this case because we are “overly impressed by the fact that if [the bystander] proceeds, he will bring it about that more live by *merely turning a trolley*”<sup>134</sup>—that is, roughly, we are more easily able to ignore, here, the (morally relevant) fact that choosing the second option would violate our negative duty not to harm, only because the method by which harm is inflicted in this case is so bloodless and sanitized.

The details of Thomson’s argument will not concern us further. What I should like to explore in this paper, rather, is the general structure of the reasoning Thomson employs here, and the understanding of normative ethics and moral theory it, as I shall argue, implicitly presupposes.

It will be useful to begin with an oddity in Thomson’s program already visible in the short précis above: Thomson writes (repeatedly) that her aim is to “explain” the verdicts we give in the various Trolley Problem permutations. One might expect her rather to have tried to justify those verdicts. In any event, she needs some contrast between her task with regard to these verdicts, and her task with regard to the aberrant verdict, *sc.*, our ‘intuition’ (as we might say) that the bystander may turn the trolley. That contrast comes out, ultimately, as one between explaining why we may do such-and-such, on the one hand, and explaining why it *seems to us* that we may do such-and-such else, on the other.<sup>135</sup> The question then becomes: what substantiates this distinction?

Before trying to answer this question, it is helpful to remind oneself that a distinction of roughly this shape is quite familiar to us: we are not yet enmeshed in controversial philosophical commitments when we say, for example, “It seemed all right to me then (to do such-and-such),

---

<sup>134</sup> *Ibid.*, p. 374; the emphasis has been added to retain the sense of the original sentence in context.

<sup>135</sup> “Turning the Trolley,” p. 374.

but I later realized that it was totally unjustified”); we are entirely familiar with a distinction between something’s being ‘all right’—justified, permissible, licit—and merely *seeming* to be all right. Of course, if such a judgment is controverted, the pressure will be on to explain, not only the appearance (the ‘seeming’) but *why* our original moral verdict was a mere seeming—why what had originally seemed to justify the verdict now falls to the ground. There are obviously many different cases here, and we shall return to them below; but the general pattern is clear enough.

What is the analogue for Thomson? It is not, in fact, easy to say. Indeed, read in isolation, Thomson’s explanation for the aberrant verdict might look very much like a justification. Thus, in explaining why people tend to give the right verdict for certain other, morally symmetrical cases, but the wrong one to Bystander, she writes,

The more drastic the means, the more strikingly abhorrent the agent’s proceeding. That, I suspect, may be due to the fact that the more drastic the means, the more striking it is that the agent who proceeds infringes a negative duty to the one.<sup>136</sup>

But the thought here is evidently that, in Bystander, the morally relevant facts are less likely to *strike us*—where ‘striking’ connotes not *a measure of wrongness* but a psychological fact: ‘strikingly’, in the first sentence, is meant, not as an intensifier (of ‘abhorrent’), but as an adverb of manner. Our question is: what is the difference between the explanation offered here, for Bystander, and the other ‘explanation’, invoking the Doctrine of Doing and Allowing (DDA), for the rest of our verdicts?

The question becomes more pressing in light of Thomson’s remark that the DDA itself requires “an account of its source.” Thus she writes:

---

<sup>136</sup> *Ibid.*

[I]t is one thing to say there is a difference in weight between positive and negative duties, and quite another to say what the source of that difference is. I know of no thoroughly convincing account of its source, and regard the need for one as among the most pressing in all of moral theory.<sup>137</sup>

Surely this need is pressing: absent such an account, we might feel entitled to ask why what has been ‘explained’ is on any different footing from what has been explained away. That is: in both cases, we have a distinction available which explains why our verdicts vary systematically; but in the former case this distinction is supposed to carry moral weight, while in the latter, it is not.

In a way, this is unfair to Thomson. The language in which the explanation for our aberrant verdict is described is incompatible with a factive judgment: to be “overly impressed” by anything—here: the means of causing death—is to be impressed by it in a way that we oughtn’t be. And indeed it may just seem obvious to us that the grisliness of the means by which someone’s death is brought about could not matter morally. On the other hand, it seemed equally obvious to most of us that the bystander could indeed turn the trolley and kill the one. Is there some sort of methodological principle to which we could appeal here which would tell us which sort of obviousness should ‘win’ (i.e., be morally decisive)? One tempting answer to this question is that, absent a general principle that substantiates the particular verdict (‘intuition’), that verdict must be given up, no matter how ‘obvious’ it may appear. But then, in the first place, it is none too clear that we have some general principle in the offing about ‘grisliness of means’. Moreover, it is hard to see why some general principle should offer any support when that principle *itself* stands in need of justification (or “an account of its source”).

---

<sup>137</sup> *Ibid.*, p. 372.

The problem could be framed as follows: philosophers face a choice, in these situations, between giving up an ‘intuition’ about one case because it conflicts with a general principle (which principle, itself, has only its summarizing capacity and, *perhaps*, intuition, going for it); and searching for a new principle which covers *all* the cases. Now on what basis might we decide to pursue one route over the other?<sup>138</sup>

4. Shelly Kagan notes, in “Thinking About Cases,”<sup>139</sup> that we give our intuitions about *particular* cases a kind of priority when doing philosophy: we are inclined to reject more general principles in favor of the more particular spontaneous verdicts that contradict them. Kagan goes on to argue that this priority is mysterious. For one thing, he writes, it is not really clear that we can draw a clean line between ‘specific’ intuitions, and those which we have about ‘general’ rules: though we tend to think of (e.g.) trolley problems as ‘specific’ or ‘concrete’, such hypotheticals are not, as Kagan notes, completely filled in.<sup>140</sup> More important, we do not have a worked out moral epistemology which would enable us to say why more *particular* verdicts should be privileged.

The point is important, since one ambition of moral theorizing has always been, as we have noted, to *overturn* particular verdicts by means of more general rules. Yet we are very resistant to doing this in practice. And this resistance, justified or not, stands in the way of the project’s ambition.

---

<sup>138</sup> I am tempted to say here that Thomson’s method would be more acceptable if she had (honestly) begun with the intuition that we may *not* turn the trolley in Bystander, and then worked back to the DDA; but in fact she is quite explicit that she was forced (with everyone else) to give up her natural intuition about the case because it doesn’t fit the general rule which covers the rest of the cases. It will be complained that these objections are, in the broad sense, *ad hominem*; but if the arguments below are sound, *ad hominem* objections are just the relevant kind.

<sup>139</sup> In Paul, Miller, and Paul, *Moral Knowledge*.

<sup>140</sup> *Ibid*, 60-3.

None of this is to say that we do, still less ought to, treat our specific moral intuitions as absolutely sacrosanct; and perhaps this is all a method like Thomson's requires: one intuition, in particular, she regards as aberrant, a view to which she was led, apparently, by the fact that it cannot be covered by a general principle—the DDA—which covers all the other similar cases (that philosophers have yet dreamt up). Of course, to hold that intuitions are not sacrosanct is not yet to violate our practice of giving a kind of *priority* to intuitions<sup>141</sup>: it is only to say that our intuitions are not infallible, and that much is demanded by any sane account, since different people have contradictory intuitions over a single case: it just can't be that *everyone's* intuitions are *all* correct.<sup>142</sup> But there are different kinds of fallibility, and they are worth distinguishing. Thomson, in effect, treats intuitions as *evidence*; and that is not an unusual position to take—Kagan himself uses this language repeatedly. Evidence is of course defeasible, but when one concludes that some piece of *prima facie* evidence for some conclusion or other does not, in fact, support that conclusion, one has an obligation to explain the original appearance; Thomson is attempting to discharge this obligation when she tells her psychological story about why people tend to be cool with pulling the switch in Bystander. But there are other ways to be wrong.

Thomson's program could be described as the floating of a moral hypothesis—and of course, she is far from alone in this; we might call it the standard method of first-order ethics. The moral hypothesis in question is the Doctrine of Doing and Allowing; and the evidence for this hypothesis consists in our various spontaneous moral judgments—our 'intuitions'—about various hypothetical cases which are consistent with this Doctrine. And just like in an ordinary

---

<sup>141</sup> Though as Kagan notes, one counterexample of the specific-intuition type is usually regarded as sufficient for annihilating a theory (see *ibid.*, p. 45). Indeed, the construction of counterexamples to general principles might with equal justice be called the 'main method' of first-order ethics: see below in the text.

<sup>142</sup> In a way, this is too strong, since it would be possible to hold, e.g., that people only *appear* to disagree about various cases and that, should everyone be properly apprised of all the 'descriptive facts', everyone would agree; I myself sincerely doubt this to be the case, and regard it as a pious hope. Alternatively, one might plump for some brand of subjectivism—equally unappealing.

case of hypothesis-confirmation, one must find some way to ‘account for’ the evidence that doesn’t fit: here, a psychological story about how we might be misled into making the aberrant judgment about Bystander.

But is there something strange about the very idea of moral hypotheses—presupposing, as it does, the idea of a moral reality so alien to us that we have to make hypotheses about its nature? One way to put our discomfort here (presuming to reader to share it) would be like this: for a putative moral rule to have any interest for us, it would have *itself* to strike us, intuitively, as correct, and this *independently* of any of its instances; otherwise, it is very hard to see why we should give up one of our spontaneous moral judgments (‘intuitions’) in favor of the rule. Is the moral realm neat? Easily codifiable? Simple and elegant? It is hard to see just why anyone should think so.<sup>143</sup> And if it is not, then it is hard to see why simplicity, elegance, and the like should count as merits in a rule that has nothing else going for it. –Which is not to say that the Doctrine of Doing and Allowing itself has no independent purchase on our moral sensibility; only to say that, *if* it does not, or if its only purchase comes by way of reflection on those intuitions that actually instantiate it, then it is hard to see why such a rule should compel us to suspend our judgment in Bystander.<sup>144</sup>

The point can be brought out by considering an extreme form of a moral hypothesis: suppose that nine-tenths of our moral intuitions in some area could be accounted for by the Rule

---

<sup>143</sup> There are some reasons, but I doubt their cogency: it might be suggested, e.g., that the moral realm must be simple enough for creatures like us to be able to understand it and make judgments about it; but in the first place, we seem to make judgments perfectly well—not to say ‘morally’, that would of course be begging the question; I mean: without difficulty—without the DDA. Moreover, it’s hard to see why, on a view which countenances the very idea of moral hypotheses, anyone should suppose that the moral reality should be understandable by men at all; perhaps it is just too complicated for us, and we are doomed to forever getting it wrong. Cf. Ch 3. See also Bernard Williams 1985, pp. 105-6 and 113 for similar doubts.

<sup>144</sup> And it is far from clear that Thomson regards DDA as intuitively compelling—she has, after all, resisted it for many years. (Cf. fn. 138, above.) Moreover, her demand for “an account of its source” is plausibly read as a demand for a story that makes the DDA intuitive. (On the other hand, it may be, what is perhaps odd in its own way, a demand for a justification for something we *already* find intuitively compelling.)

of Open Doors: “Do that which maximizes the opening of doors.” I mean: suppose that, as it turns out, applying this rule in any particular case of action will give a verdict that is, nine-tenths of the time, in accord with our pre-theoretical moral intuition. No one (I hope) would dream of supposing that we ought to overturn the remaining one-tenth of our moral judgments on the strength of this elegant rule. The example is extreme, but the point is general: no mere inductive hypothesis which accounts for a good deal of the ‘data’ should convince us to throw out the remaining, ‘aberrant’ moral judgments—nor, I take it, by the same token, could it be thought to *justify* those intuitions with which it is consonant.

The example may be thought a cheat, insofar as it imports our skepticism about our moral judgments ever being summarized by *such* a rule, into our intuitions about whether rules that are only as plausible as their instances might have some ‘normative force’ against aberrant (by the lights of the rule) judgments. That is, if it strikes us as implausible that the Rule of Open Doors might ever have normative force against our moral intuitions, this (it may be argued) is only because it strikes us as implausible that this rule ever really *could* summarize (nine-tenths of) our moral judgments. For this reason I am tempted to point to such rules as, “Do that which maximizes the propagation of your genes,” or “Do that which will result in the greatest happiness for the greatest possible number”; such rules probably do cover anyway a good deal of our ordinary moral intuitions; but unfortunately, people have gone so far as to actually believe these rules, so I suppose they will hardly help my case.

What we need, in any event, is a closer investigation of the idea of moral hypotheses, and of the alternatives, with a view to answering this question: what sort of normative force—that is, force against our moral intuitions—might a general moral hypothesis have?

5. It seems to me there are two broad forms that a moral hypothesis might take. On the one hand, if we decide to take quite seriously the analogy so often put forth between scientific theorizing and moral theorizing, we may think of the generation of moral hypotheses as, in effect, an exercise in curve fitting: our intuitions are the data points, the general principle describes the curve. Such general moral rules could then turn out to be, upon their discovery, wholly novel or unthought of—at the extreme, the Rule of Open Doors. On the other hand, one might suggest (as Kagan does) that the general rules are themselves *familiar* rules, which have some native intuitive force. (I shall try to bring out the character of this second option more clearly in what follows.)

Between these two poles, there are mixed options. We might have, for example, a rule which, though relatively novel-sounding, has the backing of some larger metaphysical apparatus (our concept of ‘persons’, for example<sup>145</sup>). But here too we can distinguish between rules which, absent the metaphysical structure, are basically unprepossessing, and on the other hand those rules which are already plausible, and for which the metaphysical apparatus is brought in just in order to ‘shore up’ the rule.

We have then a kind of continuum, from rules which strike us as capturing some moral truth, to rules that seem to us no more attractive than the Rule of Open Doors: mere summaries, as we might say, of our reactions thus far. My concern in this section is only with rules at the latter end of the spectrum. These might indeed be called the pure form of a ‘moral hypothesis’. The question is why we should allow any such rule to overwrite a first-order intuition. On the

---

<sup>145</sup> Cf. F.M. Kamm 2007, p. 5; 1992, pp. 6-11. Kamm is unusual among moral theorists, as she recognizes, in her reluctance to overwrite a first-order intuition; her view is that, in the face of intuitions which contradict some rule, we should simply make the rule more complicated, in order to account for that intuition. By my lights, this is superior to the alternative. But Kamm’s view that the discovery, and subsequent justification of, such general principles is necessary (see again 2007, p. 5), I find rather doubtful; I come to this point in §9.

face of it, it is hard to see what reason there could be; it would seem to amount to (what Smart in another context called) ‘rule-fetishism’. If the point was to capture, in a general rule, the data of our first-order intuitions, then it would seem clear enough that any (robust) contradictory intuition simply falsifies the general principle; it will have to be junked, or emended.<sup>146</sup>

It is sometimes said in this context<sup>147</sup> that in discovering such rules, we are discovering facts about the ‘deep structure’ of our *psychology*, or principles which we “unconsciously” hold—i.e., some psychological *mechanism*. The assumption, explicit or implicit, that there *must* be some such mechanism, is widespread.<sup>148</sup> But whether or not any such mechanism ‘must’ be there, it would seem to be the grossest commission of the ‘naturalistic fallacy’ to then suppose that we ought to overturn first-order intuitions which conflict with some (preliminary?) statement of the rule governing that mechanism. That I *tend* to make judgments in ‘accordance’ with some general rule could not show, by itself, that I *ought* to make such judgments; still less could the fact that I *usually* make judgments in accordance with it mean that I should *always* do so.

*A propos* of these remarks, we might also note (as I did above) the pervasiveness of the language of ‘evidence’ in connection with this mode of moral theorizing.<sup>149</sup> Now the term ‘evidence’ can be used in all sorts of ways, of course; but in common parlance, it is used to strike up a contrast between *direct* and *indirect* apprehension of some fact.<sup>150</sup> This fits a picture on which our moral intuitions are regarded as indirect ways of coming to know general principles; given *enough* evidence, we are justified in ‘believing’ the general rule. –That is, in believing

---

<sup>146</sup> We could compare here the Unified Theory of Jokes. Surely, if the goal is to discover the ‘rules’ governing what we find funny, a sound methodology will discard, or anyway amend, rules which have run into counter-examples.

<sup>147</sup> For instance, by Kamm, *ibid*, at fn.4, pp. 5 and 8.

<sup>148</sup> McKeever and Ridge 1996 are explicit: see pp. 147-8. The antidote to such a view is found in McDowell 1979 and 1981. My own arguments against such a view are given in Chapter 3.

<sup>149</sup> Kagan, *ibid*, is typical.

<sup>150</sup> Cf. Austin’s “Other Minds,” reprinted in his 1961, on the contrast between “How do you know?” and “Why do you believe?”; he associates the latter with *evidence*, and that in turn with ‘indirect knowing’: see pp. 78-82.

that the general rule represents a genuine normative fact—albeit one that has been operating, as it were, behind our backs.<sup>151</sup>

The burden of Chapter 1 was to argue that such a view leads to intolerable skeptical puzzles.<sup>152</sup> I suggested there an alternative model for moral knowledge, one on which we only come to see *that* things are so *in light of* the reasons *why* they are so. (And hence one where we query a claim by asking “Why?” rather than “How do you know?”) Such a model precludes the possibility of obtaining moral knowledge on the basis of evidence; and hence precludes moral hypotheses: the fact that many of our judgments are fit by some particular ‘general rule’ would go no way towards giving ‘evidence’ that such a rule was ‘true’. If it were evidence for anything, it would be evidence that our judgments were the upshot of certain psychological mechanisms; or that they had their origin in (say) certain pressures from our evolutionary past—hardly grounds for overturning new, recalcitrant first-order intuitions.

Similar considerations were unearthed in the discussion of moral testimony (Chapter 2). My remarks there centered on the idea of *thick moral concept possession*; and I suggested that testimony was not guaranteed to transmit moral knowledge because it could not, by itself, guarantee the transmission of the relevant concept. But the point might plausibly be broadened. My intermediate conclusion there was that ‘Humean grounds’—i.e., evidence, in my terms—were not sufficient to generate spontaneous (moral) judgments. In place of our aesthetic analogy, we might consider jokes again. Could knowledge of *just which jokes are funny* be

---

<sup>151</sup> I suppose it is idle speculation, but I have sometimes wondered whether the actual model for the method of moral hypothesis was not the investigations of natural science after all, but rather something more like the search for laws that have been promulgated by some legislator, and are only now dimly apprehended by us. In that case we really could have something like a dispassionate investigation; the results of that investigation might surprise us; and we could get something like *evidence* that various laws were ‘in effect’. Such a view of course raises familiar worries: we might wonder how the law ‘applies’ to us; or why we should ‘care about it’.

<sup>152</sup> For an argument that has some parallels to my own, see Setiya (nd); his concern is with skepticism stemming from considerations about *disagreement*.

passed on by testimony? Well, certainly I could *tell* someone which jokes were funny (supposing I knew)—and even ‘the reasons why’. This might enable her to, e.g., tell someone else; or even: fake a laugh. But if she is utterly humorless in general, then I think we shall say that all my testimony has not given her—as we could put it—‘comedic knowledge’: she will not be making any ‘spontaneous comedic judgments’.

By the same token, a method of moral hypothesis—or a Unified Theory of Jokes, for that matter—could not transmit any spontaneous moral judgments; and therefore it could not be a method of obtaining moral knowledge. In each case, the reason is the same: the conclusions reached are reached on the basis of *evidence*, or ‘Humean’ grounds; and this falls short of producing the sort of ‘seeing that things are so *in light* of the reasons why’ that is necessary, if I am right, for genuine moral knowledge. –Which is not to say, of course, that someone might not be set straight morally by being confronted with some general rules—or set straight by receiving some ‘testimony’. But in each case, the Humean grounds—the evidence—will drop away as irrelevant: the fact that my interlocutor is sincere in asserting *p*; or that *p* fits a curve described by some general rule; these will play no part in my coming to see that *p* is so.

These points come out with more force if we could be allowed to suppose, what I did for dialectical purposes in Chapter 2, that moral knowledge is centrally a kind of conceptual knowledge or know-how, *sc.*, knowledge of how to deploy various ‘thick’ moral terms. What the method of moral hypothesis presupposes is that it is possible to learn that, e.g., some particular action is wrong, without coming to *see* that action as wrong. In the present terms: it supposes that we could come to learn part of the extension of the concept of (e.g.) *justice*, without having achieved genuine mastery of that concept—without being able to apply it and reapply it in novel situations. Like moral testimony, the method of hypothesis could only

generate a kind of counterfeit ‘knowledge’—the ability, at best, to fake someone else out on occasion, about the extent of our moral knowledge.

6. But perhaps the idea of a ‘pure’ moral hypothesis had nothing going for it in the first place. After all, the general moral rules that are offered in these projects are always supposed to have *some* intuitive weight.<sup>153</sup> The question—raised, e.g., by Kagan—is why we should give more weight to *particular* moral judgments than to these general rules which are, if not always familiar, anyway never as alien as the ‘Rule of Open Doors’.

Before turning to that matter, however, I should like to begin by offering a sketchy and inevitably inadequate defense of the thesis that has come to be known as ‘judgment’ or ‘motivational’ internalism. This is the view, roughly, that there is some sort of necessary connection between any given moral judgment, and some motivation: someone who believes that it is wrong to  $\Phi$  will be at least ‘partially motivated’ not to  $\Phi$ —and so on.

The thesis remains extremely controversial;<sup>154</sup> and I cannot hope to deal satisfactorily with it here. But a few words on the topic are in order, and I should like to offer a mild defense.

In the first place, motivational internalism comes in many flavors; I mean only to assume, as I say, a very weak form of the thesis, namely:

---

<sup>153</sup> It might be noted in this connection that the methods of moral theorizing are inevitably ‘dialectical’: it is a presupposition of such works that the reader *shares* the writer’s intuitions about the particular local cases deployed; the effort is always to move the reader from *some* things she believes to some *other* things—and not merely to retail alien facts, as we might find, e.g., in a treatise on chemistry.

<sup>154</sup> For a recent review of the literature, see Björklund, *et al.*, 2012; they note that moral internalism is “one of the most debated theses in contemporary metaethics” (125). The classic defense is probably Smith 1994, especially 71-96 (he defends there a weak form of the thesis which he calls the ‘practicality requirement’). Other defenders of motivational internalism are various as: Blackburn 1998 (see pp. 61-5); McDowell 1978, 1979; Stevenson 1937 (see p.16); and Price 1787 (especially at Ch. 8, p. 194). Deniers of the thesis include Copp 1997 and Shafer-Landau 2003.

WEAK MORAL INTERNALISM: if someone judges that she morally ought to  $\Phi$ , then she is at least partially motivated to  $\Phi$ , unless she is in special defeating circumstances.

This formula is consistent with someone believing that she ought to  $\Phi$  and yet, for all sorts of reasons, failing to do it: weakness of will, ‘depression’, etc. It may even be consistent with some forms of the ‘amoralist’<sup>155</sup> argument.

It will nevertheless be objected, as it has been repeatedly in the literature, that we can conceive without contradiction of someone—a complete amoralist—who has not now and never has had any motivation to act in any way morally, and yet for all that believes that various things are right and wrong.

I think it is worth noting here the striking parallel between this internalist/externalist debate, and the classic 20<sup>th</sup> century debate in the philosophy of mind over whether pain is intrinsically connected to ‘pain behavior’. Here too, skeptics have offered images of stoics in great pain who nevertheless betrayed nothing; of whole races of ‘super-Spartans’ whose pain need *never* be displayed overtly; and all sorts of other thought experiments by which these theorists sought to sever the supposedly conceptual or anyway *a priori* tie between pain and its expression.

My own view is that these two strands of thought—skepticism about the connection between pain and pain behavior, and motivational externalism—contain a common error, namely, an inability to see how a connection can be both necessary and, at the same time, extremely elastic.<sup>156</sup> This point can be brought out by considering a classic objection to the idea

---

<sup>155</sup> For one of many versions of the amoralist argument against internalism, see Shafer-Landau 2003, p. 146ff.

<sup>156</sup> Blackburn 1998 emphasizes the importance of a ‘background’ of normal conditions—people being motivated by their moral judgments—against which we can make sense of the amoralist. But the idea that, e.g., the ‘normal’ situations must *outnumber* the abnormal ones is probably a mistake. Perhaps we need here some kind of ‘non-Fregean generality’: see Thompson, 2008, especially the Introduction, and Part One, Chapter 4.

of moral internalism, when it is married to moral *cognitivism*. “How,” it is asked, “could the mere *belief* that something ought to be done motivate one to *do* it?” This question is supposed to produce a dilemma: either we drop cognitivism, and allow that moral judgments are a kind of desire; or we drop motivational internalism, and allow that someone—anyone—can have some belief about what he ought to do without being moved to do it.

The objection tends to take on a metaphysical character: how could some *belief*-state have the property of being intrinsically motivating? But surely the correct question is not about the properties of mental states, but about *when we attribute moral beliefs* to people. And there is nothing mysterious in the idea that one criterion for the attribution of a *moral* belief to someone—some other or even, perhaps, oneself—is that she be partially motivated to act in compliance with it. Here the familiar examples from the internalist literature come to mind: if someone consistently failed to do as she claimed that she ought, we would very commonly doubt either her sincerity, or her self-knowledge. Indeed, without some such criterion, the idea of attributing moral ‘beliefs’ to people would seem to become something like pointless—anyway, would involve cutting the link between such beliefs and *action*.<sup>157</sup>

And yet for all that, the connection between motivation and belief can, as I say, be very elastic: if *in general* someone tends to do as she says she ought; and moreover, *this* case is one that (say) only *comes up* when she is already placed in circumstances of great temptation; in such a case it would be possible to nevertheless credit her with the belief which she does not ever actually act on, but still avers. What is important here, however, is that *we have to suppose such*

---

<sup>157</sup> Cynically, then, one might say that moral philosophy had moved on from a concern with the moral *agent*, to a concern with the moral *critic* or *judge*, as Stuart Hampshire (1949, p. 467) once complained; and then from concern with the *critic* to concern with some kind of mere *knower*: moral knowledge as cut off from action and even from praise and censure.

*things as that she is in a circumstance of great temptation.*<sup>158</sup> So that the question as to the truth of motivational internalism is not, I think, as has often been supposed, merely whether it is possible to imagine someone who believes she should  $\Phi$  and yet does not; but also and at the same time what *else* we have to imagine in so imagining her.<sup>159</sup>

These points will of course not be enough to persuade the determined externalist. But I would emphasize here two points. In the first place, and perhaps rather vulgarly, the fact that internalism still remains a live option means that, even if my points in what follows go through only for those of an internalist persuasion, something anyway will have been shown; and in the second, I am not sure that what I say below really does require a commitment to internalism. For even externalists of all stripes largely believe that there is *some* kind of deep connection between moral judgments or beliefs and motivation; what they deny is only that it is a *necessary* connection; and it may be that a ‘deep’ connection is enough.<sup>160</sup>

7. If Weak Motivational Internalism, or something like it, is true, it places a constraint on—indeed, if what I have said above is correct, it *is* a constraint on—our attribution of moral beliefs to individuals; and this will have certain ramifications for moral hypotheses. In particular, we could not attribute to someone a moral belief merely on the basis of her *avowing* it—indeed, not

---

<sup>158</sup> Cf. Wittgenstein 1953, §391, on pain and pain behavior: “And it is important that I have to imagine an artful concealment here.”

<sup>159</sup> Parallel remarks might perhaps be made about the ‘complete’ amoralist, but I am not so sure. My own inclination is to say that, insofar as I can imagine such a person, I imagine him to be either utterly insincere, or only voicing his moral ‘beliefs’ in a familiar, Harean, inverted-commas sense (see Hare 1952, §7.4). On the other hand, it is very hard to know how much one’s intuitions have been damaged by theory; for some empirical evidence that non-philosophers are by and large externalists, see Shaun Nichols 2002. For a kind of ‘attributionist’ internalist proposal in something like the same spirit as my own, see Jon Tresan, 2006.

<sup>160</sup> See e.g. Peter Railton 1986; and Sigrún Svavarsdóttir 1999. Each admit as real, and thus try to give externalist explanations for, the deep connection between judging one ought to  $\Phi$  and being motivated to  $\Phi$ . Additional support for something like motivational internalism might come from consideration pertaining to moral *testimony*: see Chapter 2 of the present work.

even on the basis of her *sincerely* avowing it; for she may just be *wrong* about what she thinks she believes, and that will be shown in her actions, and the moral judgments she makes about others, in cases real and hypothetical.

If something like this is right, then it is clear enough in the first place, I think, when we might attribute to someone (what we can call) a *particular* moral belief—e.g., that it would be wrong for her, now, to do *this*; or, slightly more generally, that it would be wrong for anyone, in such-and-such a situation, to do *this*. Thus anyone who finds herself with the ‘intuition’ that one may *not* (say) push a fat man in front of a trolley to prevent the death of five others, evidently believes that it would be *wrong* to do so. And then, if the foregoing was on the right track, we can say: it is correct to attribute to someone such a belief—*sc.*, that it would be wrong to push the fat man—on the grounds that she is *partially motivated* by that belief. Partially motivated how? Here, the motivation expresses itself in disapproval; hopefully, it would manifest itself in *action* by a refusal to push the fat man in a real situation. And likewise for any other number of ‘particular’ beliefs—beliefs, that is, about particular situations.

We can now ask when, on the other hand, we might correctly attribute to someone the belief in some *general rule*.

Here it will be useful to distinguish between two kinds of general rules—what we might call ‘thick’ and ‘thin’ rules. ‘Thick’ rules will be those that employ some thick moral term, like ‘honesty’ or ‘courage’; ‘thin’ moral rules, on the other hand, we can describe as ‘if-then’ statements, with some non-moral content on the left (say, ‘this is a lie’) and a thin term on the right (‘this is wrong’). But either sort of rule could be compressed into a command or an ought-statement: “Don’t lie,” say, or “One ought not to lie”—as opposed to “Be honest,” or “One ought to be honest.” It may be that these two rule-kinds mark out the ends of a continuum, as opposed

to flat alternatives, but then keeping the two poles in view will in any case be enough to orient us.

Now it is a notorious fact that very simple thin rules—like “It’s wrong to lie”—are inevitably inadequate. Indeed it is just their inadequacy which has helped inspire the moral theorizing project, and its attempt to devise ever more elaborate rules which can capture our moral intuitions. But it is an equally notorious fact that *thick* rules have struck many theorists as hopelessly vague—as insufficiently determinate to deliver the kind of answers we want from a moral theory.<sup>161</sup> But one further, and related point to note about these rules is the differential ease with which we are able to attribute them to other people.

Thus suppose I (naively) attribute to Jones the rule “don’t lie”: I take it, on the basis of his conduct, that this is one of his moral rules. Now if I should discover Jones generally telling lies in one sort of situation; or approving of them in thought experiments; and so on; then I shall have to emend the rule: “don’t lie—unless.” Whereas if I had begun by attributing to him the rule, “Be honest,” then such conduct need not undermine nor alter the attribution of this rule, but only alter my understanding of his application of the term ‘honesty’.

None of this is to say of course that, by applying the latter method, I shall find myself in *more agreement* with Jones than otherwise; it is only that the disagreements that we have, when we have them, will be located in a different place: we shall disagree, not over what the correct moral rule is, but over some of the applications of the concept ‘honesty’.

---

<sup>161</sup> I suppose it might be argued that “Be honest” is as implausible a moral rule as “It’s wrong to lie”: should we really always be honest? There is a classic answer to this problem, which answer, however, exacerbates the other problem: on one understanding of the thesis of the unity of virtues, it is impossible to be honest and behave badly at the same time; thus where there is an apparent conflict between (say) honesty and kindness—such examples are easy enough to dream up—and the correct thing to do is (say) to be kind, then, this view tells us, in being kind we are not after all being *dishonest*. And contrariwise, when honesty compels us to tell someone something that will hurt her, we are not being unkind in so telling her; indeed, in such a situation, it would be, not only dishonest, but no kindness to hide the truth. Cf. McDowell 1979.

What is interesting is that it appears that the same will be true of ascriptions of particular and general moral beliefs to *myself*. I can discover my own *particular* moral beliefs by being confronted with (say) novel trolley cases: by eliciting my own moral intuitions. And I can attribute *general* moral beliefs to myself on the basis of (say) the patterns of these responses. But, if the foregoing was sound, then it will be, in general, easier to attribute to myself—on the basis of my own conduct, and my patterns of approval and disapproval—certain ‘thick’ rules than their ‘thin’ counterparts. –Or rather: the ‘thinner’ the rule, the more likely it will be that I shall later have to retract this self-ascription.

8. Now Kagan considers the “deflationary” possibility<sup>162</sup> that our deference to our intuitions about *particular* cases rests merely on the fact that we are more *confident* in them. But if what I have written above is on the right track, then our deference to particular intuitions over more general ones rests not on our greater confidence that the former are *true*, but on our greater confidence, as we might put it, that we ourselves *actually believe them*.

There will no trouble, in general, with attributing to myself the belief in certain thick moral rules: such rules will not be contradicted by any new particular verdict; though I shall, in the course of discovering such verdicts, discover the contours of certain thick concepts of mine—honesty or kindness, say.

But as we noted before, the relative ease with which one might attribute to himself belief in thick moral rules is the logical complement of the relative inability of such rules to override any particular judgment: new intuitions will, in general, neither contradict nor be contradicted by such rules. Whereas it is just the opposite with the thin rules: they will more easily be found to

---

<sup>162</sup> “Thinking about Cases,” p. 46.

contradict—or to be contradicted by—novel particular moral verdicts. But the difficulty is, what I noted above, that since, in doing any bit of moral theorizing, we are always trying to figure out, not just the moral facts, but what we actually ourselves believe, every intuition which conflicts with the general thin rule will be so much evidence that we never did believe that rule in the first place.

9. This explains, I think, and justifies, our reluctance to allow general moral rules (of the thin kind) to overwrite particular spontaneous moral verdicts. But don't we *need* some sort of justification, in the form of more general principles, or some kind of general metaphysical considerations? The point of Chapter 3 was to discourage such thoughts. From the point of view of the present discussion, such principles would always come either too soon or too late—too soon, if they remained sufficiently un-vetted that they might overturn some centrally held spontaneous moral verdict (an intuition); too late, if questions at the ground level are already settled. After all, it is well known that no such general principles will convince just anyone else; and they are after all superfluous for ourselves.<sup>163</sup>

What remains possible here, I think, is the hope of some kind of more profound self-understanding that might be granted by the discovery of such principles. Sometimes such things are said. Here the details about the form of such general principles will matter. But we can prescind from such details; the general idea will be a system of general principles which, although they will not be in the business of *overturning* any particular moral judgments, will

---

<sup>163</sup> Hence I am profoundly suspicious that the idea of the “separateness of persons” is really fitted to do any work in moral theory; it is, rather, always tacked on at the end by anti-consequentialists, much too late to be taken seriously as the font of any anti-consequentialist intuitions; and much too soon to persuade any real consequentialists.

nevertheless explain or justify those judgments—perhaps the central ones—on which we all agree. And then this would enhance our self-understanding.

Perhaps. But a very mild acquaintance with history will suffice to remind us that our ancestors—indeed, our grandparents—did not share a great deal of our moral judgments. I take it as fairly obvious that our descendants—indeed, our grandchildren—will differ with us on a fair number more. I suspect myself that they will regard us as barbarous and benighted. I do not think these facts give us any reason to lose confidence in our own moral judgments. But they surely make the project of erecting a superstructure atop—or, altering the metaphor, a foundation beneath—our own considered moral judgments (supposing even that ‘we’ all agreed on them) a bit silly: our descendants will reject that superstructure root and branch. At best, they will regard it as rather quaint (as we perhaps regard certain intramural religious disputes of the last millennium); at worst, as a colossal waste of intellectual energy.

On the other hand, these same descendants will be less likely to reject such a superstructure, *the less it has to say* about particular moral judgments. At the limit, this ‘superstructure’ will have *no* implications for first-order moral judgments. Those who took the moral theorizing project seriously may now wonder what the point would be. As far as *their* project goes, I think they would be right to wonder: *this* kind of investigation into the ‘foundation’ of ethics would give no aid or comfort to the project of “a moral ‘theory’ which systematizes and extends a body of moral judgments.” But it would be a mistake to simply reject such an investigation, on these grounds; after all, given the vague terms in which I have described it, to reject *such* a project might be to reject all (or almost all) meta-ethical theorizing—surely a mistake.

**10.** The implications of all this are serious.

In the first place, it would transpire that there is something seriously doubtful about advancing some general moral principle, when that principle conflicts with an intuition that one, oneself, has. It cannot be right to say that the Doctrine of Doing and Allowing *must* be right, because it takes care of all known cases save one. On the contrary, what Bystander shows us is that the DDA cannot be correct.

—Or rather, we have no special reason to believe that it is. For all I have said here, Thomson may be right, and it may be that the ‘bloodlessness’ of the Bystander case is truly deforming our judgments. But, if the foregoing is sound, *she* is not in a position to argue that we do wrong in turning the trolley in Bystander; *she* does not even *believe* it. Only someone who *saw* turning the trolley in Bystander *as* wrong, in light of the reasons *why*, would be able to argue such a thing to us; otherwise, the supposition that it is wrong is really no more than a moral hypothesis.

None of this is to deny that we really could just decide to alter our behavior, and bring it in line with some principle unearthed in the course of moral theorizing. Perhaps Professor Thomson will no longer flip switches in Bystander cases. My thought is that this would be something like fraudulent—like forcing a laugh at jokes one doesn’t get, because the Theory says it’s funny; or proselytizing for films one doesn’t actually think are any good.

Nor could we deny that there is such a thing as authentically changing one’s mind in moral matters—and that, even, in the face of general moral principles. It might be felt that without the help of theory, we are doomed to a sort of moral conservatism.<sup>164</sup> But the fact is that there is such a thing as changing someone’s mind in ethical matters, without deploying the

---

<sup>164</sup> See the various essays in Part II of Simpson and Clarke 1989.

sophisticated, evidential methods of moral theory. Sometimes it is enough to remind someone of a general requirement of virtue. Often, of course, it is not enough; but in such cases no amount of *theory* is going to help, either—that, at least, is my experience.

Relatedly, it has been argued, for example by Robert Louden,<sup>165</sup> that moral theory—or “theory-like elements”—have been essential to moral progress. It is an empirical question, I suppose; I rather doubt it—though in fairness I am less sanguine, maybe, that the ‘progress’ has been all in one direction. But if we suspend our verdict on the last point; and moreover suppose that great moral change (one way or the other) really has been brought about by the deployment of “theory-like elements”; then one upshot of this essay is that such change may have been brought about by something like fraudulence.

For parallel reasons I am skeptical of the idea of deploying the results of moral theorizing to decide our actions in novel cases, or cases about which we *have* no real view. It would be a method, I suppose; and, in the absence of any other, perhaps just as well. But if the main lines of this essay are correct, it would amount to a more or less arbitrary way of deciding what to do; and would certainly not give us any moral *knowledge*. A grand theory of film criticism might be similarly deployed to tell us about the relative merits of movies of which we have no strong opinion; but no one would think, I take it, that we might thereby discover ‘hidden’ masterpieces.

But there are more general implications. We are all familiar with a method, found not just in philosophy texts but in ordinary life, of trying to get our interlocutor to agree to some general moral principle; and then showing him, by means of that principle, that he must give up some first-order belief. We are probably all equally familiar with the phenomenon in which our stubborn interlocutor, rather than admitting error, now simply denies that he subscribes to that

---

<sup>165</sup> Louden 1992, pp. 149-50.

principle. But if the foregoing is correct, then our interlocutor's behavior is perfectly justified: we have exactly shown to him that he never really did believe in such a principle. And it is we, in deploying such methods against him, who are engaged in sophistry.

*Pittsburgh, August 2012*

## BIBLIOGRAPHY

- Anscombe, G.E.M. 1958a. "Modern Moral Philosophy," *Philosophy*, 33(124): 1-19.
- Anscombe, G.E.M. 1958b. "On Brute Facts," *Analysis*, 18(3): 69-72.
- Anscombe, G.E.M. 1963. *Intention*, Ithaca, NY: Cornell University Press.
- Anscombe, G.E.M. 1981. "Authority in Morals," in *Ethics, Religion, and Politics: Collected Philosophical Papers Volume III*, Minneapolis: University of Minnesota Press.
- Austin, J.L. 1946., "Other Minds," in *Philosophical Papers*, 1961, Oxford: Clarendon Press.
- Baier, A. 1989. "Doing Without Moral Theory" in *Anti-Theory in Ethics and Moral Conservatism*, S. Clarke and E. Simpson, eds., Albany: State University of New York Press.
- Berker, S. 2009. "The Normative Insignificance of Neuroscience," *Philosophy & Public Affairs*, 37(4): 293-329.
- Björklund, F., Björnsson, G., Eriksson, J., Olinder, R.F., and Strandberg, C. 2012. "Recent Work on Motivational Internalism," *Analysis* 72(1): 124-137.
- Blackburn, S. 1998. *Ruling Passions*, New York: Oxford University Press.
- Clarke, S.G., and Simpson, E., eds. 1989. *Anti-Theory in Ethics and Moral Conservatism*, Albany: State University of New York Press.
- Copp, D. 1997. "Belief, Reason, and Motivation: Michael Smith's The Moral Problem," *Ethics* 108(1): 33-54.
- Dancy, J. 2004. *Ethics Without Principles*, Oxford: Clarendon Press.
- Driver, J. 2006. "Autonomy and the Asymmetry Problem for Moral Expertise," *Philosophical Studies* 128(3): 619-644.
- Enoch, D. 2010. "The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It," *Philosophical Studies*, 148: 413-438.

- Foot, P. 1978. "Are Moral Considerations Overriding," in *Virtues and Vices*, Oxford University Press.
- Field, H. 1989. *Realism, Mathematics and Modality*, New York: Basil Blackwell.
- Hampshire, S. 1949. "Fallacies in Moral Philosophy," *Mind* 58(232): 466-482.
- Hare, R. M. 1952. *The Language of Morals*, Oxford: Clarendon Press.
- Harman, G. 1977. *The Nature of Morality*, New York: Oxford University Press.
- Hills, A. 2009. "Moral Testimony and Moral Epistemology," *Ethics*, 120(1): 94-127.
- Hobbes, T. 1994. *Leviathan*, Indianapolis: Hackett.
- Hopkins, R. 2007. "What is Wrong With Moral Testimony," *Philosophy and Phenomenological Research* 74(3): 611-634.
- Hurka, T. 2004. "Normative Ethics: Back to the Future," in *The Future for Philosophy*, B. Leiter ed., Oxford University Press.
- Jones, K. 1999. "Second-Hand Moral Knowledge," *The Journal of Philosophy* 96(2): 55-78.
- Joyce, R. 2006. *The Evolution of Morality*, Cambridge: MIT Press.
- Kagan, S. 2001. "Thinking About Cases," in *Moral Knowledge*, E.F. Paul, F.D. Miller, and J. Paul, eds., Cambridge University Press.
- Kamm, F.M. 1992. *Creation and Abortion: A Study in Moral and Legal Philosophy*, New York: Oxford University Press.
- Kamm, F.M. 2007. *Intricate Ethics*, New York: Oxford University Press.
- Lavin, D. 2004. "Practical Reason and the Possibility of Error," *Ethics*, 114(3): 424-457.
- Leiter, B. 2004. "The Hermeneutics of Suspicion: Recovering Marx, Nietzsche, and Freud," in *The Future for Philosophy*, B. Leiter ed. Oxford University Press.
- Louden, R.B. 1992. *Morality and Moral Theory: A Reappraisal and Reaffirmation*, New York: Oxford University Press
- McDowell, J. 1978. "Are Moral Requirement Hypothetical Imperatives?" reprinted in *Mind, Value, and Reality*, 2001, Cambridge: Harvard University Press.
- McDowell, J. 1979 "Virtue and Reason," reprinted in *Mind, Value, and Reality*, 2001,

- Cambridge: Harvard University Press.
- McDowell, J. 1981. "Non-Cognitivism and Rule-Following," reprinted in *Mind, Value, and Reality*, 2001, Cambridge: Harvard University Press.
- McKeever, S.D. and Ridge, M. 2006. *Principled Ethics: Generalism as a Regulative Ideal*, New York: Oxford University Press.
- McNaughton, D. and Rawlings, P. 2000. "Unprincipled Ethics," in *Moral Particularism*, B. Hooker and M. Little, eds., Oxford University Press.
- Mill, J. S. 1978, *On Liberty*, Indianapolis: Hackett.
- Nichols, S. 2002. "On the Genealogy of Norms: A Case for the Role of Emotion in Cultural Evolution," *Philosophy of Science* 69(2): 234-55.
- Nickel, P. 2001. "Moral Testimony and Its Authority," *Ethical Theory and Moral Practice* 4(3): 253-266.
- Oddie, G. 2005. *Value, Reality and Desire*. Oxford University Press.
- Parfit, D. n.d. *On What Matters*. (unpublished m.s.)
- Price, R. 1787/1974 *Review of the Principal Questions in Morals*, ed. D.D. Raphael, Oxford: Clarendon Press.
- Quinn, W. 1989. "Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing," *Philosophy and Public Affairs* 18(4) 334-351.
- Railton, P. 1986. "Moral Realism," *Philosophical Review* 95(2): 163-207.
- Scanlon, T.M. 1998. *What Do We Owe to Each Other?*, Cambridge: Harvard University Press.
- Schafer, K. 2010. "Evolution and Normative Skepticism," *Australasian Journal of Philosophy*, 88(4): 471-88.
- Setiya, K. n.d. "Does Moral Theory Corrupt the Youth." (unpublished m.s.)
- Shafer-Landau, R. 2003. *Moral Realism: A Defence*, New York: Oxford University Press.
- Sher, G. 2001. "But I Could Be Wrong," *Social Philosophy and Policy*, 18: 64-78.
- Singer, P. 2005. "Ethics and Intuitions," *The Journal of Ethics*, 9(3): 331-352.
- Smith, M.1995. *The Moral Problem*, Oxford: Basil Blackwell.

- Stevenson, C.L. 1937. "The Emotive Meaning of Ethical Terms," *Mind* 46(181): 14-31.
- Street, S. 2006. "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies*, 127: 109-66.
- Street, S. n.d., "Objectivity and Truth: You'd Better Re-Think It," (unpublished m.s.)
- Stuchlik, J. n.d. "Connatural Ethical Knowledge." (unpublished m.s.)
- Svavarsdóttir, S. 1999 "Moral Cognitivism and Motivation," *Philosophical Review* 108(2): 161-219.
- Thompson, M. 2008. *Life and Action*, Cambridge: Harvard University Press.
- Thomson, J.J. 2008. "Turning the Trolley," *Philosophy and Public Affairs* 36(4): 359-74.
- Timmons, M. 1999. *Morality Without Foundations*, New York: Oxford University Press.
- Tresan, J. 2006. "De Dicto Internalist Cognitivism" *Noûs* 40(1): 143-65.
- Urmson, J.O. 1974. "A Defence of Intuitionism," *Proceedings of the Aristotelian Society*, 75:111-119.
- van Fraassen, B. 1980. *The Scientific Image*, New York: Oxford University Press.
- Wedgwood, R. 1999. "The Price of Non-Reductive Moral Realism," *Ethical Theory and Moral Practice*, 2(3): 199-215.
- Wellman, C. 1971. *Challenge and Response: Justification in Ethics*, Carbondale: Southern Illinois University Press
- White, R. 2010. "You Just Believe That Because..." *Philosophical Perspectives*, 24(1): 573-615.
- Williams, B. 1979. "Internal and External Reasons," reprinted in *Moral Luck*, 1981, Cambridge University Press, 101-13.
- Williams, B. 1985. *Ethics and the Limits of Philosophy*, Cambridge: Harvard University Press.
- Williams, B. 2002. *Truth and Truthfulness: An Essay in Genealogy*, Princeton University Press.
- Wittgenstein, L. 1953. *Philosophical Investigations*, trans. G.E.M. Anscombe, Oxford: Basil Blackwell
- Wittgenstein, L. 1967. *Lectures and Conversations on Aesthetics, Psychology, and Religious Belief*; Cecil Barrett, Ed., Berkeley: University of California Press..