Moral Knowledge and the Genealogy of Error

1 INTRODUCTION

The study of human history can be unsettling, not in the least because of our ancestors' seemingly boundless capacity for inventing, refining and implementing practices which appear utterly horrific to us. History seems to present an endless litany of torture, war, slavery, cruelty and oppression, and while we may feel that this is a history of moral *error*, further reflection can provoke an unease that is more distinctly philosophical in nature. For it is not easy to say how so many people throughout human history remained completely ignorant of moral truths which seem nearly self-evident to us.

I believe that this unease reflects our tacit awareness of an epistemological problem which has not, to my knowledge, received any attention in contemporary meta-ethics. The problem arises for anyone who (1) has moral beliefs, (2) thinks that those beliefs constitute knowledge, and (3) acknowledges that a large number of mature human beings have held contrary beliefs. Most philosophers who meet this description, realist and anti-realist alike, have accepted that they must explain how their beliefs are sensitive to the moral truth. As I will argue, there is in fact another demand that such believers must meet. They must provide a *theory of error*—a contrastive explanation for how their moral opponents have gone wrong where they went right. I will show that we cannot claim to have vindicated our own moral reliability unless we can explain the unreliability of those who hold contrary beliefs. This, in turn, requires us to engage directly with cultural history, a topic which has been unfortunately obscured by the meta-ethicist's near-exclusive focus on evolutionary challenges to moral belief.

While this challenge theoretically applies to anyone with moral beliefs, it is highly probable that any reader of this paper will be committed to a few basic liberal-Enlightenment values concerning racial and sexual equality, the importance of self-determination and the moral irrelevance of social class. Since I share those beliefs, I will often write as though it is *their* vindication which is at stake here. In doing so, I do not mean to imply that only those who hold these beliefs need a theory of error. Any believer needs such a theory in the face of manifest moral diversity.

In the opening sections, I'll motivate the epistemological importance of theories of error. My aim will be fairly ambitious: starting from widely accepted principles, I will try to prove that the inability to explain widespread error means that we have no explanation for our own reliability. I'll then spend the bulk of the paper evaluating (and rejecting) various theories of error that have been offered, either explicitly or implicitly, by those who have defended moral knowledge. I will not reject such theories because they are necessarily false or because they suffer from any deep internal confusion. Rather, I will show that they simply cannot be made to shoulder this particular explanatory burden, and that we must therefore search for an alternative if we are to vindicate our right to claim that we have moral knowledge.

2 THE NEED FOR A THEORY OF ERROR

2.1 THE RELIABILITY CHALLENGE

Reliability challenges are now a familiar feature of the meta-ethical landscape.[1] The fundamental idea behind such challenges is easy to state. A believer is reliable if they tend to get things right, if the processes which are responsible for their beliefs are significantly likely to produce true beliefs. In other words, if their beliefs could only be *coincidentally* true, this makes their beliefs unreliable or untrustworthy.[2] The reliability challenge asks us to show, at bare minimum, that our beliefs were likely to be true, given the processes and methods which are responsible for them. Of course, it would be too demanding to ask for conclusive *proof* that our beliefs are by and large likely to be true. Rather, we seem to face a more reasonable burden, one that is met if we provide a realistic explanation for our most central moral beliefs which strongly suggests that most of those beliefs will be true. Absent such an account, we do not need to draw the strong conclusion that we *are* unreliable, but it is clear that our beliefs will have been fairly seriously undermined, and that we may not have the right to say that they qualify as knowledge.

I should say a few words about how I am construing this challenge. As most moral philosophers know, Gilbert Harman articulated a similar epistemological challenge

[1] See Sharon Street, "A Darwinian Dilemma for Realist Theories of Value," *Philosophical Studies,* Vol. 127, No.1 (2006): 109-166; David Enoch, "The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope With It." Philosophical Studies, Vol 148, No. 3 (2010): 413-438; Benjamin James Fraser, "Evolutionary Debunking Arguments and the Reliability of Moral Cognition," *Philosophical Studies* Vol. 168 No.2 (2014): 457-473; Joshua Schechter, "Explanatory Challenges in Metaethics." In T. McPherson and D. Plunkett (eds), *Routledge Handbook of Metaethics* (New York, Routledge, forthcoming).
[2] See Duncan Pritchard, "Sensitivity, Safety, and Anti-luck Epistemology," In J. Greco (ed), *The Oxford Handbook of Skepticism.* (Oxford, Oxford University Press, 2008), pp.437-455.

which asked us to show how our moral beliefs (or 'observations') are explained by moral facts. Subsequent discussions refined this challenge, with some arguing that it is met if we can show that moral facts explain *anything whatsoever*, for example, a single moral belief, a social institution or an action.[3] However, I take it that this result would not vindicate our moral reliability, since it would not show that most of our central moral beliefs are likely to be true. The reliability challenge, then, is in an important sense more demanding than Harman's challenge. However, this, I take it, does not diminish its importance: if I acquire positive reasons to think that I am generally unreliable in some domain, this might very easily undercut my justification for a great many of my beliefs in that domain.

Next, it is important to bear in mind that *every* non-skeptic needs to explain their own reliability, even an anti-realist who grounds domain-specific facts in contingent, subjective mental states. Moral subjectivists like Sharon Street simply think that they have an easy time doing so, especially by comparison with the moral realist, who has to explain how our cognitive faculties track mind-independent facts.[4] Yet, this claim is not actually as straightforward as it seems, since a subjectivist vindication of our reliability might be bought at an extremely high price. Street might have to admit that many historical moralizers have been equally reliable, even when they have believed that slavery was just, that women were fit to be ruled, and so on. We should not lose sight of the fact that this is a very high price to pay, and as we will see, Street herself tries to avoid paying it.

The reliability challenge has provoked a huge array of responses, each of which tries to show, in broad outlines, how human populations might be reasonably expected to acquire at least some true moral beliefs, *ceteris paribus*. In my view, such responses must meet a subsidiary requirement, namely, they must provide a theory of error. I'll now try to show why this is so.

2.2 THEORIES OF ERROR

What exactly is a theory of error? Fundamentally, it is an *explanation* of an error, and not the identification of one. Suppose a child believes that a stick is bent, and I believe that it is straight. I am committed to the idea that the child is mistaken. But to give a theory of error is to go beyond this and *explain* the child's mistake. For example, I might note that the stick is immersed in water, that in such circumstances

---

[3] See Nicholas Sturgeon,. "Harman on Moral Explanations of Natural Facts." *Southern Journal of Philosophy* Vol 24 (1986): 69-78; Geoff Sayre-McCord, "Moral Theory and Explanatory Impotence" in Sayre-McCord (ed) *Essays on Moral Realism*. (Cornell, Cornell University Press,1988): 256-281.

[4] Street, Ibid., and , Matthew Bedke, "Intuitive Non-Naturalism Meets Cosmic Coincidence," *Pacific Philosophical Quarterly* Vol. 90, No. 2 (2009): 188-209.

human visual systems produce perceptual impressions of bent sticks, and that the child has not had sufficient experience of such situations to correct the belief inductively. The explanation does not entail that the child is wrong: it is not meant to, since it concerns the reliability or trustworthiness of the child's belief, and not its truth. It is meant to explain how the error occurred, under the working assumption that it did.

At first blush, it can seem as though the vindication of our own reliability does not require a theory of error. If we develop theories of competence which explain our generally getting it right in some domain, haven't we thereby vindicated our beliefs? Moreover, can't we simply assume that error will be explained by the *absence* of whatever processes or factors explain our beliefs?

Unfortunately, these thoughts are mistaken. Imagine Suzy, who believes P. When asked to explain why P counts as knowledge, she cites a certain process—a general, truth-finding mechanism which is in fact shared by all mature humans. So far, so good. However, suppose she discovers a large group of mature adult humans who believe not-P. She is stuck with what I will call the *problem of explanatory symmetry*: if her explanation for how she got things right applies equally well to the belief-formation processes of those who got it wrong, she no longer has a good explanation for how she got it right.

The lesson here is straightforward. When there is widespread disagreement in some domain, a complete response to the reliability challenge must take the form of a *contrastive explanation*.[5] We need an explanation which shows why true belief obtained in one set of cases rather than in some other set.

Moreover, the numbers matter, here, since in order to trigger this requirement, the disagreement needs to be fairly widespread. Suppose Suzy passes by an aquarium and forms the belief that a stick in the aquarium is bent. If asked why her belief is reliable, she will probably cite the reliability of the human visual system, which she used to form the belief. In addition, suppose she encounters 1000 normally sighted people who, like her, believe that the stick is bent, and only one who thinks that it is straight. This in no way calls her vindicatory explanation into question, since any psychological mechanism is bound to produce the occasional false belief, even if it is normally very reliable. The dissenter, then, poses no threat to her explanation, which cites a very general truth-finding mechanism. But if she goes on to discover 10,000 normally sighted dissenters, a distinct explanation for their belief is plainly required. Suzy could not merely continue to explain her reliability by reference to the human

---

5 Bas Van Fraassen, *The Scientific Image* (Oxford, Oxford University Press, 1980).

visual system, for that very system has just been shown to produce beliefs which are contrary to hers. It is only when she discovers some *further* fact which explains the disagreement that she has a chance of explaining her own reliability.

I hold this position not merely because it is intuitive, but because it follows from widely-accepted principles that all participants in this debate ought to share. Uncontroversially, good explanations of contingent facts render their *explananda* significantly more probable than contextually salient alternative outcomes, and when Suzy discovers that her explanation cites a mechanism which is at least as likely to produce contrary beliefs, she has therefore discovered that her explanation, as it stands, is inadequate. It follows that successful vindications of our reliability, in domains containing a significant amount of disagreement, must discover asymmetries between the processes which produce our beliefs and those which produce the apparently false ones. This, in turn, requires us to explain the false ones, or to give a theory of error.

In addition, not every explanatory asymmetry will give us a satisfactory theory. The reliability challenge, recall, requires us to show that *true* belief is likely to emerge, given our best account of the historical processes responsible for our beliefs. However, there are often many equally reliable processes in a given domain. If we try to explain two sets of beliefs by respectively citing the influence of each of these processes, we will not have explained how one set was likely to contain true beliefs while the other was likely to contain false ones. In other words, explanatory asymmetry is necessary but not sufficient for a theory of error. The explanatory asymmetries must be epistemically *relevant* such that they show the set of true beliefs to be acquired in a significantly more responsible or reliable fashion.

2.3   META-ETHICAL IMPLICATIONS

There are two important implications of what has just been said. First, since a theory of error is only required when there is widespread, fundamental disagreement in some domain, the moral domain looks far more problematic than the domains of mathematics and logic. As I will soon argue, there are very deep and widespread historical differences in moral belief, and there is simply no analogous disagreement in mathematics or logic. It is very hard to find cultures who (a) possess the relevant concepts and (b) have large swaths of false logical or mathematical beliefs.[6]

---

[6] This is denied by Justin Clarke-Doane. See Justin Clarke-Doane, "Morality and Mathematics: The Evolutionary Challenge," *Ethics*, Vol 122 No.2 (2012): pp. 313-340. However, Clarke-Doane focuses unhelpfully on disagreements amongst those who theorize *about* mathematics, and does not, to my mind, show that there are fundamental mathematical disagreements amongst people who actually deploy mathematics as part of ordinary practice. It is not common for engineering projects to grind to a halt because the engineers disagree over algebraic or trigonometric facts. But this is the kind of

Moreover, even if there are occasional dissenters, I've just argued that such rare cases should not be particularly troubling. Therefore, philosophers who draw upon mathematics or logic in order to employ so-called "companions in guilt" arguments cannot do so in response to my argument.[7]

Second, and relatedly, if my analysis is correct, the reliability challenge and the argument from disagreement are not separate meta-ethical problems, to which defenders of moral knowledge may devote distinct, unrelated chapters or papers. Where there is widespread disagreement in a domain, the reliability challenge becomes more demanding, requiring a theory of error as well as a theory of success. Put differently, the meta-ethical argument from disagreement can be given its most forceful articulation if we refuse to follow many philosophers in re-casting it as a question about the resolvability of disputes under *ideal* conditions. Anyone who thinks they have moral knowledge must be able to make a principled distinction between the actual explanation for their own beliefs and the explanation for contrary beliefs, and this is *not* inquiry into what idealized disputants would come to believe. I am suggesting that we retain the focus which was originally suggested by Mackie, who noted that non-skeptics must satisfactorily explain how most past societies could have "seriously inadequate and badly distorted" moral beliefs and perceptions.[8]

So much, then, for the epistemological significance of error. I'll now try to show how the messy, disturbing details of human history can make it very difficult to explain how so many people have got things so wrong for so long.

3 THE RELIABILITY CHALLENGE AND HUMANITY'S ILLIBERAL PAST

It is natural to think that two stages are involved in any attempt to show that an agent's beliefs 'track the truth'. The first stage involves giving a theory of truth for the type of proposition in question, and the second involves showing how our beliefs were likely to be true, given the processes which have produced them. In the moral domain, these requirements are met by (a) the provision of something like a moral theory, and (b) an explanation for our beliefs which shows them to be

---

disagreement that would be analogous to the moral case, since moral disagreements prevent ordinary people from realizing shared projects of all kinds.

[7] See Russ Shafer-Landau, *Moral Realism: A Defence.* (Oxford, Oxford University Press, 2003); Hallvard Lillehammer, *Companions in Guilt: Arguments for Ethical Objectivity.* (London, Palgrave Macmillan UK, 2007).

[8] See John Mackie, *Ethics: Inventing Right and Wrong,* (New York, Penguin, 1977); ch.1.

sensitive to the truthmakers named in (a). This will constitute an explanatory vindication of the beliefs in question.

At this stage of the dialectic, we may distinguish broadly between *causal* and *rationalist* modes of explanation. Causal theories are (understandably) offered by philosophers who think that moral facts are natural facts, whereas the non-causal, rationalist mode of explanation is favored by those who deny that moral facts are identical to or reducible to natural ones. However, as I will now show, these explanations can make human history appear somewhat puzzling.

To begin with the naturalists, they claim that moral facts can (and, in favorable circumstances, do) causally influence human moral beliefs. Moreover, the dominant tendency is to argue that human societies are regulated by processes which ensure that there is a significant (if not overwhelming) tendency for groups of mistaken moral believers to be replaced by groups who possess true beliefs.[9] In most cases, this involves the suggestion that immorality (and the false belief which accompanies it) is severely disruptive to human groups, such that groups with false moral beliefs are unstable and likely to be replaced by more stable groups with beliefs that are closer to the truth. This, in a nutshell, is the naturalist's explanation of our reliability.

Now, under the assumption that something like this is true, human moral history does present the naturalist with something of a puzzle. Given that this selective pressure exists, how did human civilizations spend roughly 10,000 years mired in false belief? History is littered with cultures displaying remarkable longevity in spite of the fact that a large majority of their members had few of their basic needs met, and were subjected to socially sanctioned brutality, threats of slavery or torture, culturally based discrimination and virulent misogyny. Given the naturalist's hypothesis, this phenomenon is surprising, and it cries out for explanation. This explanation will constitute a theory of error, an explanation of how our historical rivals went wrong.

Non-naturalists, on the other hand, often place explanatory emphasis on our capacity to *reason*.[10] They admit that previous human societies may not have possessed a great deal of moral knowledge. But once human beings began—perhaps haltingly and with a great deal of difficulty—to apply their capacity for rationality to the moral domain,

---

[9] See Peter Railton, "Moral Realism," *Philosophical Review* Vol. 95 No.2 (1986): pp. 163-207; David Copp, "Darwinian Skepticism about Moral Realism," *Philosophical Issues* Vol 18 No. 1 (2008): pp. 186-206.

[10] Derek Paarfit, *On What Matters*, (Oxford, Oxford University Press, 2011): pp. 534-42; Peter Singer, *The Expanding Circle: Ethics, Evolution, and Moral Progress*, (Princeton, Princeton University Press, 2011); William FitzPatrick, "Debunking Evolutionary Debunking of Ethical Realism." *Philosophical Studies* Vol 172 No.4 (2015): pp. 883-904.

true beliefs began to emerge. Our beliefs are, it is suggested, partly the product of this long, laborious process, and this is how we are reliable.

However, this hypothesis can also make human history appear somewhat puzzling. As we will see, much depends on the precise characterization of rationality provided by the noncausalist, but such a characterization cannot be so stringent as to preclude past moralizers from possessing a faculty of practical reason. How, then, did whole populations of human reasoners, blessed with basically the same faculties as you or I, manage to miss various fundamental moral truths for so long? Why *was* this process so laborious? As is the case with the naturalist, this fact is not inexplicable in principle, but it certainly cries out for an explanation, one which must indicate some explanatory asymmetry between us and our historical ancestors.

Now, before proceeding directly to an examination of existing theories of error, I should briefly mention two strategies I won't discuss very much in this paper. Each of these strategies entails that no theory of error is needed, because no errors were actually made.

3.1   THE CONCEPTUAL RELATIVISM STRATEGY

The first, available to those who think that basic moral truths are *analytic* truths, involves the claim that we do not share any (or most) of our basic moral concepts with past societies.[11] For those theorists, the content of a concept is fixed by certain basic, platitudinous judgments made using that concept. Divergence in such judgments entails divergence in concepts, and this means that moralizers who make radically different judgments are literally talking past one another. This strategy faces a barrage of objections, not the least of which concerns our apparent ability to *agree* with past moralizers. The analytic theorist who takes this route must deny that we can do so, and one can expect, for example, neo-Aristotelians to balk at the thought that they cannot really agree with anything that Aristotle said in the *Nicomachean Ethics*. That said, I have already stressed that the epistemological challenge in this paper is a problem for those who think we *do* disagree with most past moral systems, and this entails, *a fortiori*, that we share a reasonably wide array of the concepts contained in those systems. Indeed, the leading proponent of the analytic view has shown a willingness to say that apparently mistaken moral systems are simply *immature* rather than operating with different concepts.[12] If this option is pursued, we need an explanation for this immaturity, for how so many societies were ignorant of the platitudes which fix the content of their own concepts.

---

[11] Frank Jackson, *From Metaphysics to Ethics*, (Oxford, Oxford University Press, 1998).
[12] Ibid. pp. 133

3.2 THE MORAL PARITY STRATEGY

The second strategy is suggested by a naturalist hypothesis described above. A defender of moral knowledge might note that past societies faced conditions which differed radically from our own, and conclude that their beliefs were largely *true,* relative to their situation. Our beliefs are true, their beliefs are also true, and thus a kind of moral parity exists between us and them.

One theory which can generate this result is moral relativism. A relativist claims that moral truths are grounded in certain contingent cultural or subjective configurations.[13] Such a theory leaves space to argue that the beliefs of most people in most past cultures are true, and enables a stalwart relativist to defuse the epistemological problem at hand by denying that most past moralizers were in error.

It is important, however, to note that every viable relativist theory places a certain distance between actual psychological or social configurations and the truth-grounding ones. Thus, in spite of a confusion that occasionally arises in the literature, relativism does not *automatically* generate moral parity by guaranteeing that every individual or every culture has true beliefs.[14] This matters because even die-hard relativists tend to worry about moral parity, since it involves the admission that some pretty awful behavior was *correctly* sanctioned by past societies. For example, when the ancient Babylonians drowned women for committing adultery, the relativist may have to say that they were being guided by true moral belief. For these reasons, both subjective and cultural relativists have sought to argue that much of this belief is in fact mistaken. Sharon Street, for example, thinks that her variety of subjectivism will actually generate intuitive, non-revisionary conclusions about the reasons that actual human beings have.*[15]* But to say this is to say that huge numbers of human beings have been mistaken about their own reasons, and this relativist owes us a theory of error.

A different way to generate moral parity is to embrace a sophisticated consequentialism, as naturalistic moral realists Peter Railton, Richard Boyd and David Copp have. According to this kind of theory, moral beliefs are true when they

---

[13] See Gilbert Harman, "Moral Relativism Defended," *The Philosophical Review* Vol. 84, no. 1 (1975): 3-22; David Wong, *Moral Relativity.* (Los Angeles, University of California Press, 1984); Sharon Street, "What is Constructivism in Ethics and Metaethics?," *Philosophy Compass* Vol 5 no. 5 (2010): pp. 363-384.

[14] After defending a form of moral relativism, Richard Posner directly infers that "no useful meaning can be given to the expression 'moral progress'… we cannot think of ourselves as being morally more advanced than head-shrinkers and cannibals and mutilators of female genitalia." See Richard Posner. "The Problematics of Moral and Legal Theory," *Harvard Law Review* (1998): pp. 1637-1717.

[15] See Sharon Street, "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters," *Philosophical Issues* Vol 19 no. 1 (2009): 273-298.

correspond to the norms which will in fact maximize overall happiness or welfare.[16] This latter theorist can simply accept that in certain harsh conditions, a system of norms sanctioning slavery, patriarchy and hostility to outsiders is the best available option. The belief "slavery is required" is therefore *true* in some cultural contexts, though it is false in ours, and moralizers inhabiting each context are for the most part getting it right. However, this strategy is unlikely to generate moral parity across the board: most historical cultures defended a husband's right to abuse his spouse, and it is not at all clear how a norm like this could possibly contribute to overall well-being. Moreover, like the relativists, most actual consequentialists will blanch at the thought that their theory generates a *duty* to enact human enslavement in a wide variety of cultural contexts, and will want to tell an empirical story according to which this duty is rarely, if ever, actually generated. Once again, this theorist will have to provide a theory of error, a theory of how people mistakenly believed that slavery was a moral necessity.

In sum, there are two ways for a defender of moral knowledge to avoid the provision of a theory of error. However, they involve (a) denying that we share moral concepts with past societies, or (b) affirming that a great deal of patriarchal, militaristic, xenophobic and inegalitarian belief is true. These options are very hard to accept, and therefore I will assume that a full defense of moral knowledge requires a theory of error.

Thus, under the assumption that a great deal of people in past societies at least appear to have disagreed with us, the overall meta-ethical dialectic looks like this:

[Figure 1.jpg]

Figure 1: Who Needs a Theory of Error?

Thus, any meta-ethicist who rejects moral skepticism, analytic conceptual relativism and moral parity must provide a theory of error. Bearing this in mind, I'll now proceed to examine and criticize existing theories. I will show that each faces at least one of two serious problems:

> SYMMETRY: The explanation for the epistemically impoverished situation of our cultural ancestors applies equally well to us.

---

[16] See Richard Boyd "How to Be a Moral Realist." In Geoff Sayre-McCord (ed) *Essays on Moral Realism*. (Cornell, Cornell University Press, 1986): pp. 181-228; David Brink, *Moral Realism and the Foundations of Ethics* (Cambridge, Cambridge University Press, 1989); David Copp, *Morality, Normativity, and Society*, (Oxford, Oxford University Press, 1995). Strictly speaking, Copp writes in terms of a society's fulfillment of its "basic needs", yet, once this notion is spelled out, it turns out to be virtually identical to other utilitarian-naturalist theories I've cited, since it involves ensuring the welfare of all group members in a broadly egalitarian spirit.

> EMPIRICAL INADEQUACY: The explanation would provide an epistemically relevant asymmetry, but it fails to explain a wide range of erroneous belief.

I'll begin with causal theories of error, proceeding to examine rationalist theories in due course.

4 THEORIES OF ERROR

4.1 NON-MORAL ERROR

In response to more general worries about disagreement, many theorists have been tempted to say that most moral disagreements can be traced to non-moral disagreements.[17] This thought can be naturally extended to explain moral error, which will be explained by the distorting influence of non-moral error. The suggestion here is that our historical ancestors didn't actually have false moral beliefs at the fundamental level, rather, their higher-order moral beliefs were unfortunately affected by false non-moral ones. While this strategy is proposed by subjectivist-relativists like Street, it is also very common amongst moral realists, who emphasize the ways in which a false picture of non-moral reality can lead to a completely erroneous moral view. Thus, in addressing our disagreement with other moralizers over the equal status of men and women, Rosalind Hursthouse writes that empirically false beliefs—about such things as the relative intellectual or rational capacities—explain much racist and sexist belief.[18] An oft-repeated and similar hypothesis is that much inegalitarian moral belief rested on false beliefs about which creatures were genuinely *human*. The implication is that all relevant parties share such fundamental moral judgments as:

> Differences in reasoning ability are grounds for differential moral treatment

and

> All human beings ought to be accorded the same basic rights and freedoms

However, the story goes, misogynists and racists simply believed, falsely, that there are essential differences in reasoning ability and that some actual humans were in fact

---

[17] See David Brink. "Moral Realism and the Sceptical Arguments from Disagreement and Queerness." *Australasian Journal of Philosophy* Vol 62 No.2 (1984): pp. 111-125; David Enoch, "How is Moral Disagreement a Problem for Realism?" *Journal of Ethics* Vol 13 No.1(2009): pp. 15-50.

[18] Rosalind Hursthouse, *On Virtue Ethics*. (Oxford, Oxford University Press, 1999).

non-human. In what follows, I'll provide several reasons to think that this strategy will be generally (if not completely) EMPIRICALLY INADEQUATE.

A familiar confusion which occasionally haunts the debate over the permissibility of abortion can help to illustrate the first error involved in this strategy. Since all sides seem to agree that it is seriously wrong to kill a person, people have been tempted to think that this debate might be resolved if the warring parties could simply agree on a definition of *personhood.* However, as is often pointed out, the concept *person* is not morally neutral. Rather, its application-conditions are significantly determined by the user's antecedent moral views, and it is a mistake to think that its content could just be filled in via some purely objective, morally neutral method. All of which simply means that disagreement over abortion is *not* traceable to a "non-moral" one over the nature of personhood, because *person* is a moral concept.

The more general lesson is that a genuinely non-moral disagreement should concern the application of an uncontroversially non-moral concept. Yet, it seems very plausible to say that concepts like *rational* and *human* often function as a moral concepts masquerading as purely factual ones. Their application-conditions shift depending on which moral views happen to dominate a person's psychology at any given time. Indeed, this is the only way to explain how so many Enlightenment philosophers could have argued for a kind of objective moral universalism and at the same time completely failed to condemn slavery—of either the sexual or racial variety. The concept 'human', in their moral-philosophical works, generally referred to humans as we now categorize them, but when the question of racial or sexual equality was on the table, that concept's application-conditions shifted dramatically.[19]

As such, it is not at all satisfying to explain past sexism or racism by citing an allegedly "non-moral" disagreement over degrees of rationality or humanity. Just as a pro-choice advocate might restrict the concept "person" so as to exclude zygotes in order to maintain the belief that early-term abortion is permissible, a person inflamed by misogyny or racism might declare that women or outsiders are "irrational" or "inhuman". Our disagreement is at bottom a moral disagreement.

The second error here becomes apparent when we recall that work in other areas of moral psychology suggests that even our genuinely factual beliefs are often post-hoc rationalizations of antecedent moral intuition. Philosophers often assume, erroneously, that non-moral beliefs explain moral beliefs, or that human beings characteristically *base* their moral beliefs on a set of non-moral beliefs and perceptions. The alternative picture was suggested by Nietzsche, and is supported by

---

[19] See Lucy Allais, "Kant's Racism." *Philosophical Papers* Vol 45 No.1-2 (2016): pp. 1-36.

much contemporary moral psychology, including Jonathan Haidt's 'dumbfounding' experiments, where subjects are found to retain confidence in their fundamental moral judgments even when articulated factual justifications are shown to be unsound.[20]

So, when a defender of moral knowledge wishes to say that some moral error is *explained* by a non-moral error, there are several difficult empirical hurdles they have to clear. Not only do they have to show that the relevant disagreement is unambuiguously non-moral, they have to avoid a subtle conflation which occasionally haunts the literature on moral disagreement. This is the conflation of *rationalization* and *explanation*. This point can be illustrated via an example. Suppose I say to you:

> (N) "Cockroaches are awful creatures, and we ought to exterminate them."

You disagree, and ask me to defend these normative beliefs. I respond:

> (F) "Well, they spread disease."

Here is what *is* true: as a purely logical matter, (F) can rationalize (N). In other words, there exists a valid argument which uses (F) as a premise and which terminates in (N). The argument looks like this:

> (N2) Anything which spreads disease is awful and ought to be exterminated.
> (F) Cockroaches spread disease.
>
> **Therefore**, cockroaches are awful and ought to be exterminated.

But the fact that this argument exists does not entail that (F) *explains* (N).[21] In fact, it is not even very good *evidence* that this explanatory relation exists. Indeed, in this particular case, that empirical hypothesis is implausible: if I am at all typical, my belief (N) is probably explained by disgust, and (F) is a post-hoc rationalization of (N). If you show me good evidence that (F) is false, I will probably cling to (N) anyway. Since this phenomenon is arguably ubiquitous in the moral domain, we should conclude that even when historical agents have articulated false factual justifications for their moral beliefs, this is only very weak evidence that those factual beliefs *explain* the moral ones. In a great many cases, the explanatory dependence will run in the other direction. Yet, a theory of error is an *explanation* of an error, and so the fact that a factual belief rationalizes an error is neither here nor there.

---

[20] Friedrich Nietzsche, *Daybreak: Thoughts on the Prejudices of Morality*. (Cambridge MA, Cambridge University Press, 1881/1997); Jonathan Haidt, "The Emotional Dog and its Rational Tail: a Social Intuitionist Approach to Moral Judgment." *Psychological Review* Vol 108 No.4 (2001): 814-829; Daniel Kahneman, *Thinking, Fast and Slow*, (New York, Farrar Strauss and Giroux, 2011).
[21] Nor does it even entail that I believe (N2), after all, I might just be saying this because I think that *you* believe (N2).

Taken together, these observations entail that the case for the hypothesis on the table—that many or most historical moral errors can be explained by non-moral error—is much weaker than has previously been recognized. Hursthouse is therefore committed to two strong claims, (1) that in denying *rationality* to women, historical sexists were using a non-moral concept, and (2) that most of these agents would be significantly likely to drop their sexist beliefs upon coming to believe that there are no significant differences between male and female rationality or intelligence. This claim sounds plausible when we think about our own case, but in fact there is little evidence that past moralizers would respond as we do.

Hursthouse's explanation for sexist belief fits nicely with a popular (and potentially very dubious) narrative concerning the scientific sources of our own egalitarian beliefs. However, few philosophers who offer theories of this sort have paused to consider that the basing of moral beliefs on detailed psychological facts is at best an eminently *modern* phenomenon. For example, histories of racism inform us that only during the 18th century did any human society begin to search for complex psychological justifications for their negative attitudes towards other cultural groups.[22] A newly discovered passion for the human sciences in the Enlightenment almost certainly made *our* moral beliefs more sensitive to detailed facts such as relative intellectual capacities. It may surprise readers to learn that no such passion appears in the history of xenophobic or patriarchal ideology. Rather, we are far more likely to find supporting beliefs with more respectable empirical credentials. For example, he most widely-cited history of medieval Islamphobia argues that it was an expression of large scale resentment and envy, produced by largely accurate beliefs concerning Islam's cultural and military success.[23] Of course, we think these are bad justifications, but their badness is not traceable to some obvious non-moral error. In light of these observations, we may be forced to admit that our disagreement with these moralizers is just plain old moral disagreement, as the initial appearances suggested all along.

I conclude that this strategy is not even close to as powerful as it needs to be, since there are several important ways in which it will trigger the EMPIRICAL INADEQUACY result. Doubtless, it could help to explain some moral error. But there is mounting evidence that it will not explain much.

---

[22] George M. Fredrickson, *Racism: A Short History* (Princeton, Princeton University Press, 2015): pp.56-63.
[23] John Tolan, *Saracens: Islam in the Medieval European Imagination*, (New York, Columbia University Press, 2002).

4.2   "DISTORTING" FACTORS

Rather than explain false moral belief in terms of false non-moral belief, some theorists have suggested that various social and psychological forces *distort* moral belief, pushing human beings towards error. Three of these allegedly distorting factors are commonly mentioned by causal theorists, and I'll describe each in turn before proceeding to a critique.

First, Boyd suggests that moral thinking is subject to "a very high level of social distortion" from class interests.[24] He notes that medieval moral doctrines (for example, the divine right of kings) seem to bear only a tenuous relation to general welfare, and seem custom-designed to enhance only the welfare of a small group of elites. His explanation here is that class interests biased elites towards their own welfare, such that they were driven to use their social power to construct ideological moral institutions.

Making a related move, David Brink suggests that an excessive focus on *self-interest* distorts moral judgment, such that agents fail to properly appreciate the correct distribution of benefits and burdens.[25] Indeed, the self-interest of historical agents is a powerful explanatory force in cultural history, and it is plausible to think that persons will be generally biased towards their own welfare in a way that negatively affects their moral judgment.[26]

Finally, many have noted the pervasive influence of organized religion upon past moral belief. Brink, for example, argues that organized religion has positively hindered moral inquiry.[27] This, too, is empirically plausible. To take one example, the maleness of the monotheistic God and the sin ascribed to Eve in Genesis have, without a doubt, contributed greatly to the potency of misogyny, and dogmatism in general is bad for progressive inquiry.

No empirically informed survey of moral history could omit the combined influence of class interest, selfish desire and religious orthodoxy, which has very often been extreme. Boyd and Brink are therefore to be lauded for providing empirically adequate explanations for much past moral belief. However, these explanations are not an escape from our problem, because they trigger the SYMMETRY trap. None of these theorists have provided any evidence which shows that our own beliefs are not just as pervasively regulated by these forces. Begin with religion: as Nietzsche never tired of pointing out, our own egalitarian convictions are also the product of

---

[24] Boyd, Ibid, pp.212.

[25] See Brink, Ibid, pp.205; see also Timothy Scanlon, *What We Owe to Each Other*, (Canbridge MA, Harvard University Press, 1998): pp. 358-59.

[26] Geoffrey Sayre-McCord, "Normative Explanations." *Philosophical Perspectives* Vol. 6 (1992): pp. 59.

[27] Brink Ibid., pp.206.

monotheistic religion.[28] Given that human beings do differ along a variety of dimensions, our contemporary commitment to the idea that those differences are morally irrelevant springs, as a matter of fairly incontestable historical fact, from our ancestors' belief in the equality of all souls before God.[29] Non-theists should avoid finding shallow comfort in the idea that they have escaped this influence; our basic moral instincts are deeply conditioned by our cultural context, regardless of whether we personally identify with the narratives which are dominant in that context.

Moreover, neither Brink nor Boyd provides any evidence that either self-interest or class-interests became less efficacious during the period we call the Enlightenment. If we want to avoid SYMMETRY, we had better be able to say that modern liberal-egalitarian beliefs are not particularly subject to these influences. This is going to be trickier than it might seem. For example, Judith Shklar's genealogy of egalitarian-liberalism describes it as an ideological response, on the part of the lower and middle-classes, to the threat of authoritarian terror.[30] If this is even a decent chunk of the real story, egalitarian-liberalism itself is significantly influenced by self-interest and class ideology.

Finally, I want to note that moral naturalists like Boyd and Brink in particular encounter a serious difficulty in claiming that such things as religion or class-conflict act as distorting forces upon moral belief. Notice that these forces are virtually human universals, since religion and class-structure pervade almost every society of which we have any detailed record. This striking fact is normally explained by claiming that religious practice and social hierarchies *stabilize* human societies in a wide range of contexts, providing them with shared benefits they could not achieve otherwise.[31] Yet, recall that these consequentialist moral naturalists define *true* moral belief as belief which corresponds to norms which have precisely this stabilizing function.[32] So, how could religious belief, which had the function of producing beneficial social cohesion, lead moral believers *away* from norms which produce

---

[28] Friedrich Nietzsche, *Beyond Good and Evil*, (Harmondsworth: Penguin Books, 1886/1973): s.202.

[29] See Waldron, Jeremy, *God, Locke, and Equality: Christian Foundations in Locke's Political Thought*. (Cambridge MA, Cambridge University Press, 2002); Brian Leiter, "The Boundaries of The Moral (and Legal) Community." *Alabama Law Review* Vol. 64 No. 3 (2013): pp.511-531.

[30] Judith Shklar, "The Liberalism of Fear." In Nanvy Rosenblum (ed), *Liberalism and the Moral Life* (Cambridge MA, Harvard University Press, 1989): pp. 21-43.

[31] Donald M. Broom, *The Evolution of Morality and Religion: A Biological Perspective*. (Cambridge: Cambridge University Press, 2004); Peter Richerson and Morten Christiansen, *Cultural Evolution: Society, Technology, Language, and Religion*, (Boston, MIT Press, 2013).

[32] It might be thought that Aristotelean naturalists can avoid this problem, which appears to only plague Utilitarians. In fact, this is not at all the case. After all, Aristotelians *ground* moral judgment in a sophisticated theory of self-interest: the virtues are dispositions of character which maximize individual flourishing. How could this theorist possibly turn around and argue that the influence of self-interest *distorts* moral judgment?

social cohesion? If anything, the converse is implied by the consequentialist-naturalist theory.

While I have not been able to canvass all of the possible distorting forces which might be cited by a causal theorist, I do believe that existing narratives simply don't do the job. At this stage in the dialectic, all we can do is wait for more proposals and see if they meet our standards. However, as time passes, the absence of such proposals will be strong evidence for the view that a naturalistic moral philosopher can't construct a good theory of moral error.

Of course, this is not the only option. Having examined and rejected the strategies normally adopted by causal theorists, I want to proceed to an appraisal of the rationalist model. As we have seen, a rationalist theory of error is fairly straightforward: false moral beliefs held by past cultures are explained by a failure to *reason* correctly. What should we make of this thought?

5   FAILURES OF RATIONALITY

Now, to say that past moral belief is explained by irrationality, one must deploy a determinate conception of rationality itself, not merely propose analogies with other putatively *a priori* domains.[33] It doesn't help much that moral reasoning could be analogous to mathematical reasoning, because, as I have already stressed, a theory of error is not required in the mathematical case. For these reasons, the realist needs to deploy an actual moral epistemology in order to substantiate a rationalist theory of error.

Fortunately, there are two available moral-epistemological models which might seem to provide the resources for a theory of error. I'll now try to show that each of these models encounters serious difficulty with respect to our problem. The first is a variant on coherentist, reflective-equilibrium models of reasoning, and the second is a form of foundationalist intuitionism.

5.1   CONSISTENCY-REASONING

At minimum, reasoning involves the elimination of contradictions within a single set of beliefs. Some theorists have tried to give an account of progressive belief-change in the moral domain by drawing upon this minimal conception. Richmond Campbell and Victor Kumar, for example, argue that two moral beliefs—that P is permissible and that Q is not—are inconsistent when the agent also believes (1) that P and Q

---

[33] Timothy Scanlon, *Being Realistic About Reasons*, (Oxford, Oxford University Press, 2014).

only differ along certain dimensions, and (2) that those dimensions are morally irrelevant. Importantly, Kumar and Campbell say, not implausibly, that this sense of moral relevance is not itself a conscious belief, rather, it is an intuitive principle, one arising from certain emotional-intuitive faculties. Peter Singer's famous drowning-child case can illustrate the process, since, Singer suggests, our refusal to aid a far-off starving child can only be based on their being far away, in light of the fact that we will clearly save a drowning child who is close by.[34] Mere proximity, we feel, is not a morally relevant consideration, and we have to revise one of our moral beliefs concerning either our duty to aid nearby children or our permission to ignore far-off children.[35]

The corresponding theory of error looks like this: past moralizers were comparatively bad consistency-reasoners, unable to see what their deepest values or principles actually entailed. Importantly, this theory is available to a wide range of meta-ethicists: if arch anti-realist Street were to accept that non-moral errors do not significantly explain false moral belief in the past, she might have to embrace something like this theory in order to avoid moral parity. This is because she is committed to the view that agents make true evaluative judgments when their relevant non-moral beliefs are true and when their judgments cohere with their deepest values and principles.

Interestingly, Kumar and Campbell do not consider a venerable objection to this line of thought: in cases of overall inconsistency, we do not *have* to revise our beliefs about P or Q. In order to achieve consistency, we might revise our conception of moral relevance. This familiar problem haunts coherentist models more generally, since mere coherence provides little direct guidance for belief-change.[36] In other words, consistency reasoning in Singer's case might simply provoke some particularly xenophobic agent into re-affirming the importance of proximity. This matters because in-group moral favoritism is the historical norm, and because group boundaries have often been drawn by reference to physical proximity. For many historical cultures, including those which are genealogically related to our own, being from "around here" *was* a morally relevant characteristic. They would not find it difficult, as we do, to claim that a far-off starving child has no moral claim on them. They will thus display all the cognitive consistency that could be required of them.

---

[34] Peter Singer, "Famine, Affluence, and Morality." *Philosophy & Public Affairs* Vol 38 No. 2 (1972): 229-243.

[35] Richmond Campbell and Victor Kumar, "Moral Reasoning on the Ground." *Ethics* Vol. 122 No.2 (2012): 273-312.

[36] See Gilbert Harman, *Change in View*, (Boston, MIT Press, 1986).

Thus, SYMMETRY looms large for this account. What evidence is there to suggest that past moralizers were *less* consistent in this sense than we are? Basic principles of interpretation tell strongly against any charge of serious incoherence, since the primary evidence for an agent's deepest motivations is their behavior, and history is littered with behavior which betrays deep and widespread commitment to putatively awful values. Charges of incoherence can only be motivated, I would suggest, by covertly transplanting liberal-Egalitarian values into the psychologies of past moralizers, and I know of no principled reason to do this.[37]

Having discussed the prospects for a broadly coherentist theory of error, I now proceed to its main theoretical rival.

5.2 INTUITIONISM

The core claim of intuitionism is that basic moral propositions are *self-evidently* true, where self-evidence is not the same thing as obviousness. It is, rather, the property possessed by propositions which are justifiably believed as soon as they are fully understood. The idea is that understanding such propositions necessarily produces a kind of intellectual *seeming*, or intuition, which justifies belief in roughly the way that perceptual seemings do.[38] This kind of intuition is, according to this tradition, sufficient for justified belief. Some intuitionists argue that the intuition itself is just the first-personal experience of our "grasping" a normative reason, and that this act of intellectual apprehension can explain many of our moral beliefs[39].

Let's assume that this kind of explanation can, in principle, explain our acceptance of certain basic moral propositions. How can an intuitionist explain the widespread *denial* of allegedly self-evident propositions?[40] As critics of intuitionism have long

---

[37] In order to motivate the incoherence of the immoral, a moral subjectivist might follow Hume in placing a universal concern for fellow human beings in the motivational structure of almost every human agent. He concludes that most agents have strong reasons to refrain from at least the morally worst actions. See David Hume, *An Enquiry Concerning the Principles of Morals*. (Oxford, Oxford University Press, 1751/2006), s.2. Bernard Williams, a kind of proto-subjectivist who was also deeply familiar with the vagaries of cultural history, accused Hume of suffering from "a somewhat terminal degree of optimism" on this score, and it is not hard to see how this complaint might be directed at a contemporary subjectivist (Williams 1999). At the very least, this strategy cannot succeed until we are presented with solid evidence for the extraordinary claim that the heart of a Kantian humanist beat secretly in the chests of xenophobes, patriarchs and slave-owners for 10,000 years. See Bernard Williams, "Seminar with Bernard Williams", *Ethical Perspectives* Vol.6 (1999): pp. 256.

[38] Phillip Stratton-Lake, Ethical Intuitionism: Re-evaluations, (Oxford, Oxford University Press, 2006); ch.1.

[39] William FitzPatrick, "Biology, Evolution, and Ethics". In Christian Miller (ed), *Continuum Companion to Ethics* (New York, Continuum Publishing, 2011): pp. 275.

[40] One thing they do not want to do is follow Terence Cuneo and Russ Shafer-Landau in declaring that basic moral propositions—such as those which prohibit recreational slaughter of other persons—"have gained near-universal endorsement across times and cultures." I can only urge that such

insisted, when an agent or group of agents fails to believe a supposedly self-evident proposition, this theorist owes us an explanation.[41] Moreover, this explanation must be consistent with the positive story the intuitionist wishes to give about the justificatory status of our own foundational normative beliefs.

In order to see how difficult this will be, consider Henry Sidgwick's egalitarian principle: the proposition that every agent's welfare is equally valuable.[42] This is an excellent test-case because it is, arguably, a key proposition over which we and our historical ancestors disagree. What explains the widespread, historically robust failure to accept this allegedly self-evident axiom? Many standard sources of error are ruled out, since they involve failures of *inference*, and basic moral principles such as this one are allegedly known non-inferentially. We cannot, for example, cite any formal fallacy committed by our ancestors. Nor can we say that there is some kind of independent evidence they are failing to consider, since the propositions are supposed to be *self-evident*.

These facts dramatically restrict the resources available for a theory of error.[43] Indeed, the intuitionist is basically left with the claim that our historical ancestors generally failed to understand the Sidgwickian axiom (and its contrary, which they presumably found deeply intuitive). This is a very tough pill to swallow, and we should not expect many historians to acquiesce here. Are we really to believe that ancient and medieval societies were largely composed of people who were incapable of understanding very simple propositions about the equality or non-equality of agential welfare? In addition, are we to believe that when they experienced certain non-egalitarian propositions as self-evident, that they really didn't understand those propositions?

These extraordinary hypotheses are, in fact, moderately popular in the intuitionist tradition. Many previous generations, according to the intuitionist W.D. Ross,

---

intuitionists return to cultural history, where they will find this optimistic hypothesis decisively refuted by such phenomena as Roman gladiatorial arenas, Aztec human sacrifices and public Inquisition-era *auto-da-fe*s. Should Cuneo and Shafer-Landau insist that many of these practices did not involve genuinely *recreational* killing, they will dramatically restrict the philosophical importance of such principles, which will leave room for all kinds of horrific behavior. See Terence Cuneo and Russ Shafer-Landau, "The Moral Fixed Points: New Directions for Moral Nonnaturalism." *Philosophical Studies* Vol. 171 No. 3(2014): pp.399-443.

[41] See Patrick Nowell-Smith, *Ethics*, (New York, Philosophical Library, 1957).

[42] See Henry Sidgwick, *Methods of Ethics*, (New York, Kaplan Publishing, 1901/1979), Katarina Lazari-Radek, Peter Singer, *The Point of View of the Universe: Sidgwick and Contemporary Ethics* (Oxford, Oxford University Press, 2016).

[43] One intuitionist who does not face this restriction is Robert Audi. Audi allows that basic moral propositions can be justified both inferentially and non-inferentially. Yet, it is hard to see how this helps, since true beliefs are now *easier* to come by, and erroneous beliefs are *more* mysterious. Moreover, if errors in inference are supposed to explain past moral error, Audi would need to deploy some non-intuitionist model of reasoning in order to explain this past error. See Robert Audi, *The Good in the Right: A Theory of Intuition and Intrinsic Value* (Princeton, Princeton University Press, 2005).

simply did not possess the requisite "moral maturity" to understand and attentively consider basic moral axioms.[44] But not only is this explanation empirically problematic, its deployment of a blatantly normative term ('maturity') means that it is probably just another way of saying that past moral agents didn't have the right moral beliefs. Absent an account of "moral maturity" which is not merely shorthand for "forming the right moral judgments", we have to conclude that this is not a theory of error. As we have seen, a theory of error by definition goes beyond the identification of an error, it *explains* the error, and it is hard to see how the charge of moral immaturity amounts to anything more than the mere identification of an error.

Finally, it might be suggested that deep reflection on Sidgwick's axiom was rare, because pervasive inegalitarianism ensured that no-one really sat down and thought carefully about it. At this point, we again meet our old friend, SYMMETRY. Even if this historical description were true—and I don't think that it is—we should note that precisely the same thing would be true of our society. Few of us really sit down and mull over inegalitarian propositions. This, of course, is because we are egalitarians. For these reasons, it is not at all clear that intuitionist epistemology has the resources for the requisite theory of error.

In sum, I do not believe that the invocation of *reason* is particularly promising, here. Those attracted to ethical non-naturalism because of its reliance on rationalist, non-causal theories of knowledge should, I think, be more worried than they have previously been about the prospects for a rationalist explanation of widespread moral error.

6   CONCLUSION

In this paper, I have sought to do two things. The first is to motivate the epistemological importance of theories of error, and to insist that the absence of a theory of error can, in certain conditions, entirely undermine our right to claim that our beliefs are reliably produced. The second is to illustrate just how formidable this problem is in the moral domain, given the dark and disturbing details of human cultural history. Any moral philosopher who thinks that they possess a body of knowledge which people in previous cultures lacked must say how such historical agents went wrong, and at present we have only an array of stories which are either empirically under-supported or which cut the legs out from under our own beliefs by producing explanatory symmetry.

---

[44] William David Ross, *The Right and the Good* (New York, Clarendon Press, 2002).

While my tone in this paper has been fairly skeptical, I will close by registering my hope that we can articulate empirically and philosophically satisfying theories of error that are consistent with the claim that we possess a wide body of moral knowledge. However, I also suspect that such theories will require major revisions to the metaphysical and epistemological frameworks under which many moral theorists operate. For example, we have seen that it is unwise to commit oneself to the naturalist view that moral truthmakers can simply be read off of socially adaptive norms. We have also seen that rarefied, intellectualist conceptions of reason are not well-suited for the task at hand. In my view, defenders of moral knowledge need a new theory.