Non-Naturalist Moral Realism and the Limits of Rational Reflection

Max Khan Hayward

Australasian Association of Philosophy

Routledge
Taylor & Francis Group

Non-Naturalist Moral Realism and the Limits of Rational Reflection

Max Khan Hayward

Bowling Green State University

ABSTRACT
This essay develops the epistemic challenge to non-naturalist moral realism. While evolutionary considerations do not support the strongest claims made by 'debunkers', they do provide the basis for an inductive argument that our moral dispositions and starting beliefs are at best partially reliable. So, we need some method for separating truth from falsity. Many non-naturalists think that rational reflection can play this role. But rational reflection cannot be expected to bring us to truth even from reasonably accurate starting points. Reflection selects views that are coherent and conflict-free, yet there is no reason to think that the non-natural moral truth must be like this. Inasmuch as we seek coherent, conflict-free, ethical viewpoints, that suggests that our goal is not non-natural truth at all.

1. Introduction

Non-naturalist realists in ethics famously face an epistemic challenge. This paper argues that they cannot escape it.

Non-natural facts, properties, or truths are supposedly non-causal. This seems to rule out the possibility that some putative quasi-perceptual moral insight or 'rational intuition' might reveal non-natural truths to us. Indeed, causal pressures that undeniably *have* had an influence in shaping our moral views and dispositions—those of biological and of cultural evolution—seem to have nothing at all to do with non-natural moral truth. Against these worries, non-naturalists argue that evolutionary forces are not the *only* influence shaping our moral views. Some of our moral judgments—those offered by philosophers—are also the products of rational reflection and scrutiny, and this is a reason to trust them.

One response to this doubles down on the 'debunking' power of evolutionary considerations. The rational methods to which non-naturalists appeal involve reasoning from a starting set of judgments. We revise our opinions by using some beliefs to evaluate others, checking for consistency between individual judgments, and searching for greater coherence and systematicity within our belief-set as a whole.[1] Such procedures

---

[1] Sometimes, 'reflective equilibrium' designates all such methods; at other times, 'reflective equilibrium' denotes a more specific process of working back and forth between particular and general judgments to bring them into line. Thus, I use 'rational reflection' as a catch-all term.

clearly cannot bring us to the truth when applied to a starting set of beliefs that is mostly false. The debunkers argue that evolutionary considerations show that our moral beliefs are, indeed, probably mostly false. So, there is no material for rational reflection to work with.

Many philosophers recently have argued that debunking arguments do not establish such a strong conclusion. I explain that this is correct—the debunking argument only succeeds if we set the bar for epistemic justification *in general* so high that we must accept scepticism across the board. Still, this does not end the debate. As I show, evolutionary and aetiological considerations can be marshalled to create a new and more modest argument. We are not forced to assume that our starting points are mostly false, but we must still conclude that they are *at best* a mixture of truth and error.

This, I argue, is all that we need in order to sustain the epistemic objection to non-naturalism, because rational reflection shouldn't be expected to guide us to truth in ethics, *even from somewhat correct starting points*. That's because there is no particular relation in which the ethical truths, as the non-naturalist portrays them, must stand to one another. There's no reason to expect that all correct ethical views will be assessed as 'good' from the perspective of other correct ethical views. Although beliefs that entail logical contradictions cannot both be true, it's possible to believe almost any two ethical principles without inferring a logical contradiction. The most common type of incoherence between ethical views is *practical conflict.* But the truth might contain conflict. Conflicts may detract from the theoretical virtues of a theory—their simplicity, elegance, systematicity, and so on. But there is no reason to think that the truth, as construed by the non-naturalist, will be simple, elegant, and systematic. Some ethicists think that the moral truth *must* have these features; but there is no non-question-begging argument for why these assumptions should be more reliable than any of the other ethical judgments that ethicists might reject or revise on the strength of them. Thus, non-naturalists should assume that rationally formed theories are no better than those derived from instinct or tradition. This is not quite scepticism, but it offers little comfort, considering how often we have judged the deliverances of instinct and tradition to be false.

But it's hard to deny that revising our ethical views to make them more coherent and conflict-free has nevertheless made them *better*. I suggest that this sense of betterness has nothing to do with *truth*, realistically construed. If we are moved to continue revising our theories to make them better in this sense, it shows that truth, realistically construed, might not be our goal after all. Rather, as expressivists, pragmatists, and constructivists suppose, our goal is non-alethic: 'better' ethical theories are those that are useful, agreeable, or rationally acceptable.

One note going forward: evolutionary arguments primarily target non-naturalist realism, but there is debate as to whether they extend to naturalist realism [Street 2006; Barkhausen 2016] or quasi-realism [Street 2011; Blackburn ms]. Assessing whether my arguments generalise goes beyond the scope of this essay. So I offer a challenge, rather than an objection. Some quasi-realist and naturalist theories resemble non-naturalism quite closely, such as Toppinen's quasi-realism [forthcoming] or 'Cornell realism'. These theories must answer two questions. Given the history of moral inquiry, why should they not accept my 'modest' argument about our moral starting points? And why think that the 'moral truth' must be coherent and conflict-free? Without answers to these questions, the suggestion remains that moral theorizing does not aim at truth in *any* sense.

2. What Do Evolutionary Arguments Really Show?

2.1 Reasoning from Error

Street argues that, since our 'basic evaluative tendencies' were shaped by evolutionary forces that had nothing to do with the non-natural moral truth, our moral judgments are probably wildly off-track. But many philosophers (for example, Parfit [2011] and Scanlon [2014]) respond that the countervailing influence of rational reflection on the formation of our considered moral judgments gives us a reason to trust the judgments. The influence of reason in the genealogy of morality is a vindicatory counterbalance to the influence of evolution.

Street denies that rational reflection has this power [2006: 124]:

> The objection gains its plausibility by suggesting that rational reflection provides some means of standing apart from our evaluative judgements, sorting through them, and gradually separating out the true ones from the false as if with the aid of some uncontaminated tool. ... If the fund of evaluative judgements with which human reflection began was thoroughly contaminated with illegitimate influence ... then the tools of rational reflection were equally contaminated. ... Reflection of this kind isn't going to get one any closer to evaluative truth, any more than sorting through contaminated materials with contaminated tools is going to get one closer to purity.

This might seem odd—how could *rationality* be a 'contaminated tool'? In fact, Street isn't trying to motivate evolutionary scepticism about rationality *in general*. It's just that rational reflection uses some of our judgments as a standpoint from which to critique others, and so cannot take us to the truth if not given true starting points as inputs. If our starting points are 'likely to be false,' then rational reflection is no more than [ibid.]

> a process of assessing evaluative judgements that are mostly off the mark in terms of others that are mostly off the mark.

Street's point is that rational reflection cannot turn muck into gold. Reflective equilibrium and other methods of rational reflection cannot lead us to truth from mostly false views. Her argument thus hinges on her having *already* established that most of our views are likely to be false.

So, Street's argument undermines rational reflection only if it can establish a strong conclusion—that we ought to believe that our moral starting points are 'mostly off the mark'. On the other hand, her target is limited—she aims to motivate scepticism *specifically* about moral truth, realistically construed. The argument does not aim to motivate domain-general scepticism. Indeed, she *presupposes* that we are justified in having beliefs about the external world, since her arguments rely on scientific claims about the genealogy of morality. So, the argument succeeds only if it establishes a strong sceptical conclusion about morality, whilst not presupposing an epistemic bar that is so high that it leads to scepticism in other domains.

2.2 The Improbable and the Inexplicable

What is Street's argument for the strong conclusion? On Enoch's [2010] influential interpretation, Street argues that the supposition that we have attained the moral truth entails something 'unbelievable', and is hence itself unbelievable. As Street claims, if the genealogy of our moral beliefs nowhere makes reference to their *truth* as an explanatory supposition (which it couldn't, since moral truth, according to non-naturalists, is

causally inefficacious), then it would be an 'unexplained coincidence' that the genealogy had nevertheless lead us to the truth. But what is doing the epistemic work here? Is she claiming that we can never be justified in believing coincidences? Or is it only when coincidences are unexplained that they are unbelievable?

As White [2010] points out, coincidences happen, and we are entitled to believe that they have happened when we have evidence that their results are instantiated. And we do seem to have such evidence. Whatever evidence we have for our actual moral beliefs is *ipso facto* evidence that we have attained the truth. Of course, this evidence is defeasible—*if we could not explain* how we came to the moral truth, then we would do better to adopt a sceptical attitude. So, the real question is that of whether an *explanation* for the coincidence is possible.

Some philosophers have supposed that a result's being a massive coincidence by itself implies that it is inexplicable. Coincidences are antecedently improbable, and these philosophers hold that explanation of P requires us to show why P was probable. For example, it seems antecedently improbable, given the starting physical conditions of the universe, that conscious life would have emerged. Thus, Nagel [2012] posits teleological laws to explain how consciousness arose. But almost all philosophers of science agree that the explanation of P *doesn't* have to show how P was antecedently probable. Explanation and prediction are asymmetrical. Unprotected sex with an HIV-positive partner doesn't make infection probable (transmission rates are low), but it does *explain* infection (see Jeffrey [1969] and Salmon [1971]). If we held that all improbable events were *ipso facto* inexplicable, and hence unbelievable, we would have to abandon huge swathes of our beliefs about the external world as being unjustified—including, no doubt, the evolutionary claims to which Street appeals.

If explaining how we gained true moral beliefs *doesn't* require showing how it was antecedently probable, the task of explanation looks easier. Drawing on Clarke-Doane [2015], we can see the very genealogies which debunkers give in their attempts to undermine moral belief as in fact *explaining* how we got to the moral truth. As Joyce [2006] has argued at length, we can give an adaptive evolutionary explanation for the emergence, in our primate ancestors, of the psychological dispositions, including altruism and mutualism, that undergird the formation of pro-social societal norms, and the tendency to respond to violations of such norms with *moralising* reactions. So long as there is a close relationship between the pro-social and the moral truth, Joyce's evolutionary story thus explains how we came to have at least some true moral beliefs.

Of course, this explanation only works if the moral truth *is* in fact closely related to the pro-social. And Street points out that the moral truth, as the realist understands it, could *in principle* have been anything. So, perhaps this is the sense in which the assumption that we believe the truth presupposes a coincidence—the aetiological mechanism that produced our current moral beliefs would only have brought us to the *truth* if the truth actually were largely as we happen to suppose it to be. If the moral truth had nothing at all to do with pro-sociality, then evolution would have lead us astray.

But how does this differ from our perceptual beliefs about the external world? We explain the emergence of reliable visual faculties by appeal to evolutionary advantage. But that only works in a world like ours, where visible properties of the environment have some relevance to survival. If the external world were very different—for example, if we were brains in vats—then the explanation wouldn't

work. Yet an explanation for the reliability of our visual faculties, based on the assumption that ours is not a vat-world, is rightly taken to vindicate our beliefs about the external world.

Of course, the analogy is not direct. An evolutionary story cannot be given for the emergence of a moral-perceptual faculty *specifically* tailored for detecting moral facts. According to non-naturalists, moral facts are not identical with any natural facts—facts about violations of pro-social norms are not *themselves* moral facts. The moral facts simply *supervene* on these facts. Moral truths *as such* are non-causal, and have no direct relevance to survival.

Rather, the point is that, in order to explain how we got true beliefs in *any* domain (that is, in order to avoid the kind of general scepticism that Street needs to avoid), we need to assume at least defeasibly that the world we are investigating is largely as we suppose it to be. And that is enough to offer an explanation for our having at least some accurate moral beliefs. If the *moral* world is anything like we think it is, there is a relationship between pro-sociality, altruism, and mutualism, and the realm of the moral considered as such. We needn't presuppose that every violation of altruism is morally bad, or that every moral fact supervenes on facts about pro-sociality, altruism, or mutualism. But if we explain the reliability of our perceptual systems by assuming that the external world is roughly as our perceptual faculties have lead us to believe, then we can likewise explain the at-least-partial reliability of our moralising reactions by assuming that we live in a world in which there is a fairly robust supervenience relationship between the pro-social and the moral. A world in which there is *no* connection between the pro-social and the moral is as distant from us as a world of vats, wires, brains, and simulation software. In *our* world, evolution *itself* explains why at least *some* of our moral beliefs are probably true.

So, Street's argument cannot generate the strong conclusion that our *moral* beliefs are probably largely false, without setting the epistemic bar so high as to entail self-defeating *domain-general* scepticism. But that strong conclusion is what Street appeals to in undermining reflective equilibrium.

2.3 Modest Moral Scepticism

But even if the debunking argument does not motivate thoroughgoing moral scepticism, the response above warrants only a limited optimism. It's explicable how the genealogy of morality could have endowed us with *some* correct evaluative dispositions. But we cannot suppose that biological and cultural evolution gave us *uniformly* on-track dispositions, or even ones that are anywhere near as reliable as our perceptual belief-formation mechanisms. That's not just because the moral truth did not casually regulate the progress of evolution. It's simply because we would expect natural selection to have also favoured certain dispositions that we take to be immoral. A predilection for global justice would have been disadvantageous for our ancestors; tendencies towards despicable nepotism would be evolutionarily advantageous. And, indeed, our ancestors don't seem to have had any commitment to global justice, and many of them accepted nepotism as legitimate. These examples are not exceptional. Many things that we take to be morally wrong would have been evolutionarily advantageous to believe, and indeed our ancestors had many wrong beliefs. Unless we can point to some countervailing force that would have brought us towards truth, we should, by induction, assume that we are in much the same position.

Of course, there must be some explanation of why our basic moralizing dispositions are now being pressed into the service of securing global justice and other such goals. There are two immediate explanations. First, evolutionary explanations need not assume that all features of an organism are directly selected-for; selection happens at the level of suites of adaptations that come and go together, but not all of which are directly advantageous and hence selected-for. The warmth of the polar bear's coat was selected for, but perhaps its weight was not. The dispositions that brought us to value global justice or self-sacrifice for non-conspecifics might simply have come along with the useful adaptations that allowed us to live in mutualistic norm-driven communities.

Second, we're not limited to *biological* evolution in explaining the emergence of our moral dispositions. Cultural evolution has also played a role. Perhaps it is this that pressed our basic pro-social dispositions into service in the promotion of goals like global justice. Indeed, cultural evolution also selects cultural adaptations, such as norms and belief-sets, at the level of suites—aspects of a culture's morality may simply have come along, packaged together with things that were directly selected-for. It is not hard to think of examples among the strange practices and accretions that attend to many traditional moralities.

But while these two points help us to explain the emergence of the moral beliefs that we do have, they don't give us any reason to think them *true*—there's no basis for attributing reliability to these mechanisms. Indeed, there is good reason to think that cultural evolution often leads to false moral beliefs. As Barkhausen [2016] has argued, cultural evolution selects for moral principles and practices that are mutually advantageous between parties, making its outputs wildly contingent. Depending on the parties and circumstances, extremely different moral principles can be favoured. Since *extreme relativism* is something that the non-naturalist realist presumably denies, she must conclude that some of these principles will be off-track. Any non-relativist looking at the array of moral views actually in currency in different societies will conclude that cultural evolution sometimes leads people astray. Without further argument, we have no reason to think—no explanation of how it could have transpired—that our society was a special exception. Likewise, the manner in which adaptations are packaged in suites is totally contingent. Selection will sometimes favour non-advantageous adaptations that are pro-moral, but sometimes—perhaps often—not.

So, we should assume that our starting points are probably riddled with error and our dispositions to make moral judgments are only partially reliable. The aetiology of morality might explain our having some true beliefs and reliable dispositions, but it cannot explain—in fact, it seems to rule out—our having uniformly or largely correct ones. We should expect to start with a pretty mixed bag. As such, whenever we take the influence of rational reflection to begin—be it now or in the distant past—we need to ask whether it could bring us (or could have brought us) to truth from a starting point containing many falsehoods, since that is what biological and cultural evolution probably handed to us.

My modest argument, unlike Street's, does not aim to undermine *all* of our moral beliefs. And it does not collapse into domain-general scepticism, since it allows us to defeasibly assume the correctness of our current beliefs. Starting with the assumption that the moral truth is largely as we suppose, we can see that the biological and cultural mechanisms that bring about our moral beliefs have frequently lead us astray in the past. By induction, we should assume that they continue to lead us astray. In the

absence of some countervailing force, we should assume that the starting point for rational reflection contains a mixture of truth and error.

2.4 Intuition, Instinct, and Plausibility

So, our starting set of beliefs, prior to the application of rational reflection, probably contains many falsehoods. However, proponents of reflective equilibrium often claim that the process should be applied not to *all* of our beliefs, but only to those that are antecedently selected as intuitively 'plausible', which count as our 'considered judgments'. Scanlon [2014: 82] claims that the

> force of the fact that we have arrived at certain judgments in reflective equilibrium depends on the substantive merits of the judgments we make along the way, in beginning with certain considered judgments and in modifying these judgments and others as we progress

But what are 'considered judgments'? Scanlon continues [ibid.: 84]:

> One thing one needs to ask, in deciding whether something that seems true should be treated as a considered judgment, is whether it has any implausible implications or presuppositions.

Given what's been said so far, should we expect intuitively 'plausible' beliefs to be more likely to be correct than are any others?

No. The arguments that show that our starting points are probably significantly erroneous are equally arguments that our intuitive assessments of plausibility are probably unreliable. The emergence of a set of reliable dispositions or a special faculty for intuitively detecting moral error seems highly implausible, given what we know about the biological and cultural genealogy of ethics. Indeed, such a capacity would be *counter*-adaptive, if, as argued, it is often adaptive to have false moral beliefs. A reliable capacity to intuitively find true beliefs more plausible than false ones would undermine the usefulness of false moral beliefs. So, beliefs that survive direct assessments of intuitive plausibility are no more likely to be true than are any others.

This should not be surprising. The judgments that a Christian fundamentalist or a mediaeval samurai would find most 'intuitively plausible' are no more likely to be correct than are any of their other beliefs. Indeed, given the greater relative importance that their outlooks place upon doctrinal observance and honour (respectively), compared to altruism and equality, I would expect the judgments they take to be most plausible to be *less* likely correct than their average beliefs. After all, for a fundamentalist, the goodness of altruism is subordinate to the revealed will of God—if the Text proscribes altruism, it must be rejected. For a samurai, altruism may similarly wait upon honour. If a fundamentalist or a samurai were only to reason from their most 'intuitively plausible' moral beliefs, they might exclude the correct altruistic judgments that they share with us.

Scanlon motivates the appeal to judgments of intuitive plausibility by an analogy to mathematics, where he claims that such judgments are both vital and reliable. However, the case of mathematics is different. Individual mathematical beliefs can come from poor-quality testimony, guesswork or faulty memory—of course, beliefs from such sources are unreliable. But we can see how beneficial it would have been, in a world like ours, to be disposed to form at least some of the mathematical beliefs that are actually true. And it's very hard to think of *false* mathematical beliefs that it would have been beneficial to believe. Given the starting assumption that the numerical facts supervene

on the physical facts in largely the way that we suppose they do, we wouldn't expect any significant number of off-track mathematical dispositions to have been adaptive. So, our antecedent selection of 'plausible' starting points in mathematics *should* be granted weight. Those that withstand direct scrutiny are more likely than average to be true, since these are presumably the products of our reliable mathematical dispositions, rather than memory, hearsay, and guesswork.

If all of this is correct, much rests on the competence of rational reflection to sort truth from falsehood in ethics—far more so than in other domains. In our beliefs about the external natural world, we have the advantage of causal-perceptual inputs. We cannot have these in ethics. And in mathematics we can explain why our antecedent assessments of plausibility might be reliable. We cannot similarly explain why intuitive assessments of plausibility in ethics would be reliable. And in neither our beliefs about the external world nor those about mathematics should we expect the influence of selective evolutionary pressures to be frequently false-hood-conducive. In ethics, we know that evolution has frequently selected for the unethical. So, in ethics there is surely significant work to be done in sorting truth from falsehood.

3. Why Rational Reflection Cannot Sort Truth from Falsity

3.1 Reflection as Internal Assessment

Can we expect rational reflection to weed out the false from the true? I claim that there is no reason to think so, if rational reflection consists in no more than the familiar methods of reasoning from a starting set of judgments—using some beliefs to assess others, checking our belief-sets as a whole for consistency, or seeking theoretical virtues like coherence and systematicity.

Start with the simplest forms of rational reflection. Street gives two distinct characterizations of reflective equilibrium. One version consists in 'assessing evaluative judgements … in terms of others' [2006: 124]. We can see how this would operate. We can use an ethical principle or judgment as an 'evaluative perspective', a lens through which to view the world. When we do so, we determine whether the things we so view are good, not in the 'all things considered' sense, but simply good *in terms of the value* through which we are viewing them. We can turn this gaze inward and see whether, from the perspective of one value through which we are looking, another value that we accept looks good. If it does not, we could revise or reject the 'bad' value.

The problem is that there's no reason to assume that all true value judgments will look good from the perspective of all other true value judgments. Seeming good from a true evaluative perspective obviously isn't a criterion of truth in regular factual contexts—all sorts of regrettable things are true in the world. But, even within the evaluative sphere, we can see instances of values that we accept as true, but that look unpleasant when viewed from other values. It's true that liberty is important, but insistence on liberty looks unsavoury when viewed from a perspective that judges in terms of equality. It's true that partiality towards our spouses and children is good, but this seems regrettable when viewed from the more universalistic perspective of justice. Of course, all true principles will seem good from the 'all things considered' perspective of someone who already knows the entire moral truth; but they needn't seem good when viewed from any more limited perspective.

3.2 *Reflection as Consistency*

Street also describes reflective equilibrium as a process whereby we 'test our evaluative judgements only by testing their consistency with our other evaluative judgements' [ibid.]. If testing for consistency just means eliminating logical contradiction, then this seems obviously truth-tracking: after all, *pace* dialetheists, contradictions cannot be true. However, assessing value judgments for logical contradiction is not straightforward. Does the view that 'we have a reason to maximise the good' *contradict* the view that 'we have a reason not to lie'? Not obviously—any two reasons can co-exist. We cannot immediately infer the non-existence of one reason from the existence of another.

Normally, we take belief in one fact to contradict belief in another fact if we can infer a contradictory proposition from the two. We might think that the two principles just stated *do* entail contradictory propositions, and hence cannot both be true. For example, someone who only believed the former principle would be a consequentialist, and so might infer, in a given situation, 'I have a reason to lie, and no reason not to lie', which of course contradicts the latter, deontic, principle that 'I have a reason not to lie.'

The problem is that practical judgments are inferentially non-monotonic. If I *only* believe the consequentialist principle, then I will make inferences that logically contradict the deontic principle. But if I accept both principles (in other words, if I believe in consequentialism with 'side constraints'), the inference is not the self-contradictory 'I have a reason to lie, and I have no reason not to lie, and I have a reason not to lie'; rather, it is 'I have a reason to lie, and a reason not to lie.' The fact that principle A and principle B would yield logically contradictory claims if I held each individually does not imply that the principles are logically contradictory; given the non-monotonicity of practical inference, I can believe both without inferring any contradiction. Thus, there is no *a priori* reason why I should reject one on the strength of the other, if my goal were to acquire the truth.[2]

There's a significant question as to whether the concept of *obligation* in itself rules out this kind of situation—that there's always *one* thing that you *ought* to do (although of course many philosophers argue for the existence of 'tragic choices'). But even if we think that outright *obligation* is always univocal, almost everyone accepts the possibility of *gradable* prescriptive claims, such as those concerning *prima facie* obligations, *pro tanto* rightness, or reasons for and against. Even if we think that I can never have an obligation to P and not-P, there doesn't seem to be anything conceptually impossible about a situation in which I have reasons to P and not to P, or where it is *pro tanto* right to P and not to P, and so on. When it comes to graded prescriptive and evaluative concepts, what might have looked like logical contradictions are simply *conflicts*.

This is not to say that we *never* have logically contradictory ethical beliefs. Sometimes, the correct interpretation of a moral principle is that it involves a *negative* existential statement about what reasons[3] exist, such that it directly contradicts another positive claim about what reasons exist. For example, some people who believe that 'lying is wrong' don't just mean that 'there are always reasons not to lie', but that 'there

---

[2] Briefly, the logic of this is that one cannot infer $\partial[A \& \sim A]$ from $[\partial A \& \partial \sim A]$ (where ∂ = 'ought' or, 'there is a reason to'). Holding the latter claim does not violate the principle that 'ought implies can' since, although I cannot perform $[A \& \sim A]$, I can perform each of A and $\sim A$. See Williams [1973: ch. 11, 1981: ch. 5].
[3] Or *prima facie* obligations, or whatever.

are never any reasons whatsoever to lie.' If many of our principles were of this strident form, then there would be contradictions, and thus truth-directed reasons to revise our views: it just cannot be true that we *never* have any reason whatsoever to lie, and that we *always* have at least some reason to maximize the good. These things will generate strict contradictions in practice.

It's worth noting how strong such claims are. Even someone who thought that it was always actually wrong to lie needn't accept that nothing ever counts in favour of lying. Most people who accept a prohibition on lying concede that there are sometimes considerations in favour of lying in certain situations—they simply insist that these are *outweighed* by the wrongness of lying. The starting beliefs with which we've been endowed by biological and cultural evolution seem to be mostly about what is a reason for what, rather than strident negative claims about what reasons never exist. So, it strikes me that simply purging our belief system of logical contradictions won't get us very far away from our starting points—which is worrying, if our starting points probably contain significant error.

It's true that, even once we have deleted the contradictions, we might still have a viewpoint that is *conflict-ridden.* We will often believe that we have reasons—maybe powerful reasons—that pull in opposite directions. But surely no one has any basis to think that the moral truth cannot contain conflicting reasons. Although conflict-ridden moral outlooks are unpleasant to live with, and hard to use in practical deliberation, there is nothing in the realistic notion of mind-independent moral truths, of objective reasons, that rules it out. Any two reasons can consistently co-exist. The moral truth might be ridden with conflicts between gradable prescriptions.

3.3 Reflection and Theoretical Virtue

Of course, many philosophers *have* sought to systematize their ethical viewpoints to avoid excessive conflict. For them, the presence of too many conflicts in a theory is a reason to abandon or revise that theory. This is not because conflicts are literally contradictory, and so not *possibly* true, but rather, as Kagan [1989] explains, because they detract from the *coherence* of the theory in a broader sense. Portraying rational reflection as a search for coherence in some sense that *goes beyond* mere logical consistency seems true to the classic description of reflective equilibrium as a process of 'working back and forth' between our particular and general judgments in order to bring them into 'equilibrium'. Indeed, the search for coherence in our ethical theories might be seen as being of a piece with the common preference that researchers in many domains have for theories that exhibit 'theoretical virtues'—theories that are simple, systematic, explanatory, and so on. Perhaps the broader notion of coherence sought in reflective equilibrium just *is* a matter of simplicity, systematicity, and explanatoriness; or perhaps it stands alongside these as a virtue of ethical theories. Either way, we can understand rational reflection as the search for coherence and other theoretical virtues in our ethical outlooks.

But why think that the virtues—coherence and whatever else—are indicators of ethical truth? Ethicists frequently appeal to coherence and other virtues in theory selection, but they are less careful to explain why they assume that the ethical truth will exhibit the virtues, if indeed this does occur. There certainly doesn't seem to be any *a priori* reason to think that the moral truth, as portrayed by the non-naturalist, needs to be coherent or systematic. It seems entirely reasonable to me to imagine that the truth will

be highly complex rather than simple [Griffin 2015] with a profusion of independent goods, requirements, and virtues, and that it will be full of conflict, [Williams 1973: ch. 11] with areas of moral indeterminacy [Scanlon 2014] and vagueness rather than pre-scriptions for every situation, and that individual and piecemeal judgments might not always be explained by deeper or more general principles. In other words, whatever coherence is, *over and above* non-contradiction, there is no reason to assume in advance that the truth will be coherent (or, alternatively, able to be represented by a coherent theory). And likewise for any other theoretical virtues. Perhaps this is just how things are.

I have not argued the that moral truth *does* contain lots of conflicts, or that it falls short of theoretical virtue in any other way. My point is simply that there is no *a priori* reason to rule this out, and hence no reason to expect any methodology of rational reflection that goes further than simply avoiding logical contradiction to bring us closer to the truth. All that it takes for an ethical proposition to be true on the realist view is for it to correspond to a moral fact, and there is no limit on which moral facts can exist. Parfit and Scanlon and others might hope that the truth is not such as to generate con-flicts. But it is hard to see what evidence they could have for this view. The views handed down to us by biological and cultural evolution manifestly *do* contain conflicts, and there is no *a priori* reason why these must indicate falsity.

If a philosopher insists that the moral truth *must* be coherent and conflict-free, she needs to offer some reason to think that *this* judgment is itself any more likely to be reliable than the array of other moral judgments that she will reject on the strength of it. If our first-order starting points, the ground-level moral beliefs to which we apply reflective reasoning, are likely to be partly true and partly false, surely we must assume the same of our meta-moral beliefs about the structure of the moral truth. If we have no reason to think that our dispositions to make moral judgments and to find certain claims intuitively plausible would be highly reliable, then I think that we have no reason to think that our dispositions to make judgments about the structure of the moral truth would be highly reliable. If the non-naturalist's epistemic challenge is to explain why rationally formed beliefs are likely to be true, it strikes me as simply begging the ques-tion if her answer assumes the truth of a contentious and unobvious belief about the structure of the moral truth.

Indeed, if our starting point includes *both* a messy assortment of conflicting, unsys-tematised, and poorly coherent atomic moral judgments, *and* the belief that the moral truth must be conflict-free, systematic, and coherent, then surely the simplest way to eradicate the contradiction is to abandon the latter belief on the strength of the former set, instead of embarking on the huge task of revising the former set to make it system-atic and conflict-free. So, if anything, the most basic form of rational reflection—contradiction-eradication—undercuts the assumptions needed to legitimate the more revisionary forms that aim for conflict-eradication and systematisation. In this spirit, Griffin [2015] inveighs against the distortion caused by a 'Newtonising' obsession with simplicity and systematicity. In a similar vein, Allen Wood [Marshall 2016] argues that many philosophers have been 'ravished by the formal beauty' of 'very elegant abstract formal theories' to end up with 'shallow' views that are 'revolting and inhuman'.

Philosophers appealing to theoretical virtue in theory-selection often support their methodology by pointing out that scientists do the same thing. It's widely accepted that, due to the underdetermination of theory by evidence, appeals to theoretical virtue are indispensable in science if scientists are to be able to select unique hypotheses for

acceptance. But it's a matter of debate in philosophy of science as to whether the virtues play this role because they indicate truth or for some other reason. Philosophers since Bacon[4] and Hume[5] have reminded us that an excessive love of systematicity, simplicity, and elegance can lead to false theorizing. Certain kinds of theories appeal to us, but this is no reason to think that the *world* must be like that. Many philosophers of science have denied that the virtues are well-correlated at all with truth. Levi [1997] argues that the virtues are actually negatively correlated with truth, but that virtues affect the balance of options in cognitive decision theory because virtue adds to the 'epistemic utility' of beliefs. And even those philosophers who argue that considerations like simplicity are at least defeasibly truth-indicating are careful to point out that this is a contingent, contextual fact, as in Sober's [2015] extensive examination of the virtue of simplicity. As Sober points out, we should only expect simplicity to guide us to truth given the presence of a variety of quite specific background assumptions, which don't always obtain. If seeking simplicity has often brought us true scientific theories, we can argue inductively that the virtues are good indicators of truth *in the sciences.* But there's absolutely no reason to infer that this induction carries over into ethics.

3.4 Coherence as a Practical Value

One final role for coherence and the other virtues remains, but it offers no succour for the non-naturalist. Parfit [2011] argued that there *has to be* a single true ethical principle that captures all of morality. He thought it would be 'a tragedy' if there were no one rule. But, as we have already seen, this argument cannot show us what is *true*—the truth can be uncomfortable and even unpleasant. As Blackburn [2011] says 'outside the charmed walls of All Souls College, there actually are tragedies.' But here lies the irony. Realists, and especially non-naturalists, have long argued that truth in ethics is independent of whatever anyone happens to think about it: to call something 'true' is to do something *over and above* endorsing, recommending, or approving of it. This is *why* they are forced to accept that an ethical claim's seeming like a tragedy from the perspective of some of our other ethical beliefs cannot count as evidence against its truth— the mere fact that we disapprove of some ethical belief has no bearing on its truth or falsity. One major realist criticism hurled at expressivists like Blackburn is that they allegedly cannot distinguish between calling an ethical claim 'true' and simply approving of it.

But perhaps the expressivist claim, that endorsing and judging true are closely connected, better represents the actual methodologies that we use in ethical theorizing. While it's dubious that the lack of a single overarching ethical principle would really constitute a tragedy, there are ways that the ethical world *could* (in the non-naturalist sense) be that would, I think, be tragic. If ethical conflict were powerful and pervasive, constantly placing incommensurable demands upon us, or if there were huge areas of life where the ethical facts offered no guidance at all, or if ethics required that *fiat iustitia* in a way that made it certain that *pereat mundus*,[6] then I think that *would* be a

---

[4] 'The human understanding, from its peculiar nature, easily supposes a greater degree of order and equality in things than it really finds' [Bacon 1620: Book 1, sec. XLV].
[5] 'that love of *simplicity* which has been the source of much false reasoning in philosophy' [Hume 1751: Second Enquiry, Appendix II, sec. 250].
[6] 'fiat iustitia, pereat mundus'—'Let justice be done, though the world should perish.'

tragedy. Furthermore, *because* I find such situations tragic, I am prepared to reject any moral theory that takes such forms. I suspect that such considerations are frequently employed in ethical theorizing—we seek ethical theories that are not tragic, unpleasant, or deeply uncongenial. If it is, then expressivists, who think that calling an ethical theory 'true' just *is* approving of it, can do a much better job of explaining the way that we theorize than do non-naturalist realists, who deny the connection.

Similarly, Scanlon argues that reflective equilibrium's systematising aim of 'finding general principles that account for one's beliefs' has 'important benefits' [2014: 84]. Certainly, theoretically virtuous ethical outlooks have benefits. Simple, coherent, mutually supportive, systematic, conflict-free theories are practically useful. They are an excellent basis for deliberation, discussion, and establishing social co-ordination. Inasmuch as our views are explanatory, we can explain ourselves to one another. Inasmuch as they are simple, we can swiftly work out what to do. Inasmuch as they are conflict-free, we can live without the exhaustion of guilt and moral anxiety. Inasmuch as they are systematic, we can unite our evaluations in disparate spheres of life. But, of course, there is nothing in the non-naturalist notion of ethical truth that makes us think that the correct ethical theory must be practically useful in this manner. It is ethical *pragmatists*, who see ethics as a 'social technology' [Kitcher 2011, 2012]—a practical tool to serve collective human goals, dependent on our interests rather than an ultimate authority at whose feet we must bow, regardless of the cost—who can explain why we should prefer useful theories. Again, I suspect that many ethical theorists *do* seek practically useful theories—if so, that suggests that pragmatism, not non-naturalism, is the metaethical picture that makes sense of our investigatory conduct.

It should not surprise us that reflective equilibrium and other methods of rational reflection make better sense against the backdrop of anti-realist, than of non-naturalist, metaethical pictures. Reflective equilibrium was introduced into ethical theorising by Rawls, who was a Kantian *constructivist,* not a realist. For Kantian constructivists, acceptability to rational reflection is itself the standard of correctness: there's no *further* question as to why coherence—or whatever rational reflection seeks—will also lead to truth. Acceptability to rational agents as such is all that there is to moral truth. Non-naturalists reject this by definition—the moral truth is independent of whatever we or anyone else thinks, or is disposed to think.

4. Conclusion

We can draw two conclusions from this.

Evolutionary and other genealogical considerations don't force the non-naturalist into extreme moral scepticism (as Street supposes they do). But they do show that any moral viewpoint that relies only on instinct and tradition is overwhelmingly likely to contain numerous errors. Since there is no reason to think that rational reflection will bring us closer to the truth, as non-naturalistically conceived, even rationally formed moral outlooks are probably also riddled with error. Rational reflection is no more reliable than tradition or instinct. This forces non-naturalists into a pessimistic anti-theory in ethics: the ethical systems of philosophers are no more likely to be true than are those of anyone else, and all are full of error.

But *if* we find the methods of rational reflection compelling—if we think that coherent virtuous theories are *better*, if we find the consideration that our views are the products of rational scrutiny *vindicating*—then we should abandon non-naturalism. We

should prefer theories that appear morally congenial, useful, and rational only if we believe some non-realist kind of view—expressivism, pragmatism, or constructivism. If we are not pessimistic anti-theorists, we should not be non-naturalist realists.[7]

ORCID

Max Khan Hayward http://orcid.org/0000-0003-4077-2673

References

Bacon, Francis 1620 (1905). *Novum Organum*, New York: P.F. Collier.

Barkhausen, Max 2016. Reductionist Moral Realism and the Contingency of Moral Evolution, *Ethics* 126/3: 662–89.

Blackburn, Simon 2011. Review of *On What Matters*, URL = http://www2.phil.cam.ac.uk/~swb24/reviews/Parfitfinal.htm

Blackburn, Simon ms. Sharon Street on The Independent Normative Truth as such, URL = http://www2.phil.cam.ac.uk/~swb24/PAPERS/Meanstreet.htm

Clarke-Doane, Justin 2015. Justification and Explanation in Mathematics and Morality, in *Oxford Studies in Metaethics, Vol. 10*, ed. Russ Shafer-Landau, Oxford: Oxford University Press: 80–103.

Enoch, David 2010. The Epistemological Challenge to Metanormative Realism: How Best to Understand It, and How to Cope with It, *Philosophical Studies* 148/3: 413–38.

Griffin, James 2015. *What Can Philosophy Contribute to Ethics?* Oxford: Oxford University Press.

Hume, David 1751 (1975). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, ed. L.A. Selby-Bigge and P.H. Nidditch, Oxford: Clarendon Press.

Jeffrey, Richard 1969. Statistical Explanation vs. Statistical Inference, in *Essays in Honor of Carl G. Hempel: A Tribute on the Occasion of His Sixty-Fifth Birthday l*, ed. Nicholas Rescher, Dordrecht: Reidel: 104–13.

Joyce, Richard 2006. *The Evolution of Morality*, Cambridge, MA: The MIT Press.

Kagan, Shelley 1989. *The Limits of Morality*, Oxford: Clarendon Press.

Kitcher, Philip 2011. *The Ethical Project*, Cambridge, MA: Harvard University Press.

Kitcher, Philip 2012. *Preludes to Pragmatism: Toward a Reconstruction of Philosophy*, New York: Oxford University Press.

Levi, Isaac 1997. *The Covenant of Reason: Rationality and the Commitments of Thought*, Cambridge: Cambridge University Press.

Marshall, Richard 2016. Kant Marx Fichte, *3:AM Magazine*, URL = http://www.3ammagazine.com/3am/kant-marx-fichte

Nagel, Thomas 2012. *Mind and Cosmos: Why the Materialist Neo-Darwinian Conception of Nature Is Almost Certainly False*, New York: Oxford University Press.

Parfit, Derek 2011. *On What Matters: Volume One*, Oxford: Oxford University Press.

Salmon, Wesley C. 1971. *Statistical Explanation and Statistical Relevance*, Pittsburgh: University of Pittsburgh Press.

Scanlon, T.M. 2014. *Being Realistic About Reasons*, Oxford: Oxford University Press.

Sober, Elliot 2015. *Ockham's Razors: A User's Manual*, Cambridge: Cambridge University Press.

Street, Sharon 2006. A Darwinian Dilemma for Realist Theories of Value, *Philosophical Studies* 127/1: 109–66.

Street, Sharon 2011. Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways, in *Oxford Studies in Metaethics, Vol. 6*, ed. Russ Shafer-Landau, Oxford: Oxford University Press: pp 1–32.

Toppinen, Teemu forthcoming. Non-Naturalism Gone Quasi, in *Oxford Studies in Metaethics*.

White, Roger 2010. You Just Believe that Because …, *Philosophical Perspectives* 24/1: 573–615.

Williams, Bernard 1973. *Problems of the Self: Philosophical Papers 1956–1972*, Cambridge: Cambridge University Press.

Williams, Bernard 1981. *Moral Luck: Philosophical Papers 1973–1980*, Cambridge: Cambridge University Press.