

PARFIT, CONVERGENCE, AND UNDERDETERMINATION

Marius Baumann

ONE ESPECIALLY PERSISTENT concern for moral realism takes its force from the observation that there are widespread and deep disagreements when it comes to moral issues. How could there be a truth of the matter in a field so pervaded by often fundamental dissent? This worry is at the root of what has been aptly called the *argument from moral disagreement*. Recently, Derek Parfit has provided us with a highly remarkable defense of moral realism against this argument. In his 2011 *On What Matters*, Parfit aims to show that the best versions of three of the most important families of moral theories—namely Kantianism, consequentialism, and contractualism—actually agree about what is right and wrong, forbidden, mandatory, or permissible. This, he argues, strengthens the case for moral realism vis-à-vis the argument from disagreement. If our best moral theories turn out to be in agreement about what is right and wrong, we have reason to be more optimistic about the prospect of truth in ethics.¹

Many commentators have already taken issue with Parfit's sometimes idiosyncratic reading of the authors he engages with and expressed doubts that he can achieve the promised convergence. I outline a new challenge that only sets in at a stage where Parfit considers his argument to be already won. As I see it, Parfit might have shown that moral theories can indeed agree in their extension, i.e., the set of deontic verdicts they yield for every particular act—past, present, and future.² However, he has not shown that the more theoretical disagreements, relating to the explanation of *why* certain acts are right or wrong, can also be settled. Instead, we are left with a case of extensionally equivalent yet theoretically incompatible theories. Interestingly enough, as Dietrich and List have recently observed, a structurally analogous situation has been discussed for some time now in the philosophy of science under the name of the *under-*

- 1 For better readability, I will often use “right or wrong” as a short form for all the deontic verdicts.
- 2 At least in this world. I discuss the modal strength of this claim below.

*determination of theory by evidence.*³ Following Pierre Duhem and W.V. Quine, philosophers of science have debated the possibility of there being rival scientific theories that can account for exactly the same evidence, but do so by giving explanations that are themselves incompatible. More remarkably still, this idea traditionally figures in arguments for *anti-realist* positions, a fact that has not yet attracted enough attention when it comes to the analogy to ethics.

My aim is to outline how we can reinterpret Parfit's findings in a similar anti-realist vein and thus question the effectiveness of his project to vindicate moral realism. Here is how I will proceed. I start by giving some background to the notions of moral realism and the argument from disagreement. I then outline Parfit's convergence argument and point out its main flaw, which is that it fails to address the remaining explanatory disagreements. Next, I turn to the philosophy of science for some inspiration. I introduce the idea of underdetermination and sketch how it figures in an anti-realist argument, which I subsequently adapt to the realm of ethics. Finally, I discuss three possible realist rejoinders, ultimately arguing that none of them does the trick for Parfit.

1. MORAL REALISM AND THE ARGUMENT FROM DISAGREEMENT

Let me set the stage for Parfit's reasoning in *On What Matters* (*OWM*) by introducing two crucial notions: *moral realism* and the *argument from disagreement*.

1.1. *Moral Realism: A Thin Definition and Two Additional Components*

Many of our moral claims convey the impression of stating truths about the world. If I say that your lying to your parents was wrong, or that the supreme principle of morality is the categorical imperative, or that an action is forbidden because it produces less total utility than an alternative, I seem to be stating propositions that purport to be true.

Moral realists take these claims at face value and hold that at least some of them are actually true. Sayre-McCord accordingly identifies two universally shared commitments of moral realism:

- a. Moral claims, when literally construed, have truth values.
- b. At least some moral claims actually are true.⁴

3 I came up with the analogy independently of Dietrich and List and was only recently made aware of their paper. To the best of my knowledge, Dietrich and List, "What Matters and How It Matters," includes the first explicit mention of the analogy.

4 Compare, not in exact wording but in spirit, Sayre-McCord, "The Many Moral Realisms" and "Moral Realism."

We can call this the *thin* definition of moral realism.⁵ It is obviously no more than the lowest common denominator of the different views that go by the name of moral realism. Most realists will want to add other components to their preferred version. Two options are especially relevant to our case. First, one might prefer a metaphysically thicker version that also entails something about the ontological status of moral facts or properties. For example, one might want to argue that the aforementioned moral claims can only be true if they refer to some moral property that exists independently of any natural properties. Yet this preference for a metaphysically committed view is not shared by all realists and, as it happens, is not shared by Parfit, who advocates for what he calls a *non-metaphysical* version of realism that entails no ontological claims about moral facts or properties, a position that has attracted much attention lately.⁶

Second, definitions of realism often comprise an epistemic component, relating to our ways of recognizing moral truths. That epistemic component comes in different strengths. Some realists rest comfortably pointing out solely that we do in principle have the required abilities to find out about some of the moral truths. Others take a stronger stance and line out to what extent we have already succeeded, or would under sufficient conditions be succeeding, in this undertaking. Parfit, it will turn out, subscribes to a rather strong variety of the epistemic component. This has crucial implications for what means are available to him in answering the challenge I am about to outline.

Thus Parfit subscribes to an epistemic but non-metaphysical understanding of moral realism. That means that the thin definition given above does *not* encode

- 5 I hesitate to call it the *minimal* definition, since that has other connotations. The quest to give a definition of realism has been rendered more complicated by the emergence of what has been called *minimalism* about truth or factuality, a position that allows non-realist-leaning philosophers to avail themselves of talk about moral facts or moral truths. People have consequently doubted whether the talk of moral truth (or facts or properties, for that matter) is any longer the distinguishing characteristic of realism; see Dreier, "Metaethics and the Problem of Creeping Minimalism," for a succinct statement of that view. Sayre-McCord attempts to exclude such positions by adding the proviso that we have to construe those claims *literally*. I put these issues aside for the rest of the paper.
- 6 See Parfit, *On What Matters*, 2:486. Other proponents of that view include Dworkin, *Justice for Hedgehogs*, and Scanlon, *Being Realistic about Reasons*. More precisely, Parfit is a realist about reasons, which is a distinct position that adds to the former definition of realism what has sometimes been called *reasons fundamentalism*, i.e., the view that we can analyze all normative notions in terms of reasons, which are themselves not further analyzable. I will neglect any complications that might arise from this specific view, on the assumption that all of the problems I am about to bring out could in principle be restated in terms of reasons. I will also neglect terminological subtleties, such as Parfit's preference for the label *cognitivism* as opposed to *realism*. If we understand realism in the sense just outlined, there should be no misunderstandings.

his complete view. However, it will prove useful to keep it in mind and think of the other components as *additions* that are characteristic of specific forms of realism. On the one hand, this will enable us to pinpoint more precisely where the problems with Parfit's line of argument lie. On the other hand, if Sayre-McCord is right, his definition picks out what is, *mutatis mutandis*, common to *all realisms*, not just the moral variety.⁷ It should thus be convenient as a working definition when comparing discussions of realism in different domains of inquiry.

1.2. *The Argument from Disagreement*

One classic argument against moral realism originates from the simple observation that, when it comes to morality, we encounter wide and strong disagreements. Considering how common and deeply divisive these disagreements often seem, realism faces a problem. Why should we expect there to be a fact of the matter in a field that is so pervaded by disagreements, as is the case in morality? Seeking to account for this, it would seem that we do better without the idea of as yet to be detected moral truths. Instead, other accounts, having to do with, e.g., the cultural or evolutionary origins of our moral beliefs, readily suggest themselves. These, or so it is suggested, do a better job of explaining our disagreements, and they go against the assumption of moral truths. Thus, the argument is commonly presented in the form of an *inference to the best explanation*.⁸

As it stands, the argument obviously needs improvement. The mere fact that people differ in their moral outlooks is an observation that is no less indisputable than it is inconclusive with regard to its metaethical upshots. A multitude of factors can explain the disagreements between people without challenging the assumption that there is after all a truth to be found. People might not know the relevant nonmoral facts or understand what moral claims are supported by those facts; they might be under some sort of distorting influence; or they might be using different concepts, thus merely talking past each other.⁹

What might make the case worse for the moral realist, though, is disagreement persisting between those people who have most thoroughly contemplated the issue. This, or so it is commonly assumed, is the case in normative ethics. Disagreement about the assessment of moral questions has continued to exist

7 See Sayre-McCord, "The Many Moral Realisms," 5.

8 As, for example, Mackie does in one of the defining modern statements of the argument, where he rather misleadingly calls it the *argument from relativity*, seemingly presupposing what would have to be proven (*Ethics*). Tersman points out that there are at least two alternative ways in which the argument from disagreement can be construed (*Moral Disagreement*, xiii). We do not have to go into this in more detail, though, since Parfit clearly does not have Tersman's alternatives in mind.

9 See Parfit, *On What Matters*, 2:552–63, for an extended discussion of these distorting factors.

between proponents of the most prominent and widely held traditions of moral theorizing up until today. Such is the way in which Parfit seems to think of the problem. He worries considerably about being in disagreement with other experts whom he considers to be his *epistemic peers*:

[People who] may be responding to the same evidence, their judgement in other cases may have been as reliable as ours, and they may not be more likely to have been misled.¹⁰

Indeed, Parfit fears that such disagreement might threaten our view that anything matters at all.¹¹ Yet, even if we tone down the rhetoric somewhat, it should be clear why this kind of disagreement is more threatening to realism, *provided* that it incorporates an epistemic component.¹² Granted that one can blame distorting influences for many disagreements between laymen, it would strain the realist's credibility to claim that we are able to detect moral truths, only to find out that even our best, most worked-out theories disagree about a considerable amount of nonperipheral verdicts.

2. PARFIT'S CONVERGENCE ARGUMENT AND THE REMAINING EXPLANATORY DISAGREEMENTS

This awkward situation is what Parfit has set out to change. In volume one of *OWM*, Parfit argues, among many other things, that the best versions of three of the most important families of moral theories arrive at the same conclusions about what matters.

2.1. *Convergence and the Vindication of Realism*

Parfit's reasoning for this surprising conclusion proceeds via a very detailed and

10 Parfit, *On What Matters*, 2:428.

11 See Parfit, *On What Matters*, 2:426–30.

12 That is on the assumption that one only needs to bother about people disagreeing if one holds that there is a way for us to find out about moral facts. However, as Tersman (*From Scepticism to Anti-Realism*) suggests, this might be too much of a concession to the realist. Proponents of the argument from disagreement can try to strengthen their case by insisting that absent the ability to come to know any moral truths, we are not entitled to think that there are any in the first place. Thus, showing that there are indeed such deep disagreements would suffice to challenge the realist position even if that position comes down to only the metaphysical component. I am sympathetic to that line of reasoning as an attempt to broaden the appeal of the argument beyond Parfit's specific form of realism. However, since it is not strictly needed for the overall argument, considering Parfit clearly does subscribe to an epistemic definition of realism, I will not incur the additional requirement of defending it.

intricate succession of arguments that I will not be able to restate in detail. The main structure is something like this: Parfit identifies, through a rigorous analysis of problems and objections, what he sees as the best versions of Kantianism, consequentialism, and contractualism. He is happy to acknowledge that his main interest is not in staying true to every detail of the original theories, but rather in searching for the best possible forms those theories could take.¹³ In a most remarkable move, Parfit then construes what he calls the *Kantian argument for rule consequentialism*. The argument is supposed to show that those principles that everyone can rationally will are simply those that, if universally accepted, would make things go best. Since the former formulation is Parfit's preferred version of Kantianism and the latter amounts to the best version of (rule) consequentialism, Kantianism therefore implies consequentialism. To top things off, the only principles that everyone can rationally accept are also highly likely to be the ones that no one can reasonably reject, granting compatibility with (the best version of) contractualism as well. Taking all those moves together, we can dub this the *convergence argument*.

The argument relies heavily on substantial views about reasons and rationality. In particular, Parfit is of the opinion that everyone has reasons to want the best outcomes as they would be seen from an impartial point of view (short: the optimific outcomes), and that these reasons are at least not decisively outweighed by non-optimific considerations.¹⁴ This is what ultimately allows him to argue that Kantians would indeed choose the same principles as consequentialists. That claim has already attracted considerable criticism.¹⁵ Much ink has also been spilled on whether Parfit's portrayal of the authors he engages with is a fair and unbiased one, and whether his argument for convergence goes through.¹⁶

However, these issues are not my concern here. Parfit is convinced that his preferred versions are not too far off to still be considered representatives of the three moral traditions, and I am willing to grant this as well as the convergence claim. How is this supposed to strengthen the case for moral realism, then? As we have seen, the argument from disagreement tries to capitalize on the fact that

13 Compare Parfit, *On What Matters*, 1:338–39 and 369–70.

14 Parfit, *On What Matters*, 1:377–79.

15 See, for example, Otsuka, "The Kantian Argument for Consequentialism," and Setiya, review of *On What Matters*.

16 Compare Herman, "A Mismatch of Methods"; Scanlon, "How I Am Not a Kantian"; and Larmore, "Morals and Metaphysics" for the former, and Ross, "Should Kantians Be Consequentialists?"; Wolf, "Hiking the Range"; and Darwall, "Agreement Matters" for the latter critique. What is most interesting is that representatives of *all three* of the major traditions have objected that their view gets misrepresented, and that convergence is only achieved because of an underlying bias toward one of the other frameworks.

we disagree about our moral verdicts all the time. But if Parfit is correct, this is not so for our best theories. The three main traditions agree when it comes to their main principles. Parfit does not tell us in detail which principles these are. Yet, based on his examples, we can assume that what is meant is a set of principles that specify the deontic status of classes of acts in certain circumstances.¹⁷ I will refer to these as the *deontic principles*. Since, very plausibly, verdicts in particular cases follow directly from such principles, the theories must also agree on the former. In other words, they turn out to be *deontically equivalent*. That is, they lead to exactly the same sets of verdicts about which particular acts are right or wrong, mandatory, allowed, or forbidden.¹⁸ Since these theories are also the best versions of our most important traditions, at least one of which most experts have thought to be true, this should make us confident that the verdicts we arrive at are also the correct ones.¹⁹ Ethics therefore turns out to be in no worse condition than many areas of inquiry where we are more confident that there is a truth of the matter. We may find much disagreement at first sight, but if we turn to our best theories, those disagreements vanish.

2.2. *The Remaining Explanatory Disagreements*

The depth and length at which Parfit follows through his argument for the convergence of moral theories is impressive. Nevertheless, I think that, even if the argument is successful, that is not enough to vindicate moral realism. To begin to understand why, let us first turn to one of the most famous passages in *OWM*. After outlining the convergence argument, Parfit ends volume one on a memorable note:

- 17 Examples Parfit discusses, and repeatedly modifies, include the *Consent Principle*—"It is wrong to treat anyone in any way to which this person could not rationally consent" (*On What Matters*, 1:181)—and the *Mere Means Principle*—"It is wrong to treat anyone merely as a means" (1:212).
- 18 When putting forward this claim, Parfit does not explicitly specify its modal strength, that is, whether equivalence holds for all worlds or just some subset. However, later comments suggest a strong reading, e.g., when he states: "Fundamental normative truths are not about how the actual world happens to be" (*On What Matters*, 2:489). He goes on to state that pain would be bad and rational beings would have reasons to achieve their rational aims in any possible world. As Nebel points out, this view is a direct upshot of Parfit's more general metaethical convictions ("A Counterexample to Parfit's Rule Consequentialism," 1). Nonnatural properties cannot be discovered by empirical means, yet if they were merely conditional, that is how they would have to be discovered on his view.
- 19 The point is aptly expressed by Ridge, who submits that "these three traditions (Kantian, contractualist, and consequentialist) would surely be on any reasonable person's short list of the most promising moral theories yet developed" ("Climb Every Mountain?" 60).

It has been widely believed that there are such deep disagreements between Kantians, Contractualists, and Consequentialists. That, I have argued, is not true. These people are climbing the same mountain on different sides.²⁰

The metaphor of the mountain has come to be seen as the epitome of Parfit's view.²¹ At the same time, it strikes me as particularly telling of the fundamental problem with his project. I take the metaphor to signify that when the different theorists reach the summit, i.e., when they have perfected their theories, they will notice that they agree on the same principles. Thus, there are indeed no disagreements remaining on this level. But the metaphor also betrays something else. It suggests that the theorists take different roads to the summit. This immediately prompts an additional question: Why do these differences not matter? Why is it not important how we get to the mountain?

To put it less metaphorically, I think that Parfit's argument falls short of its aim because what has been shown is, at most, that different theories can indeed lead to the same deontic principles and verdicts. Parfit thinks that he has therefore settled all the relevant conflicts. But this, or so I will argue, is not so. Though deontic equivalence might have been proven, there remain differences when it comes to those parts of the theories that go beyond the mere production of deontic verdicts.

This point has recently been stressed by Suikkanen, who accords with Parfit that the different traditions (he only considers Kantianism and consequentialism) can agree about right and wrong actions. However, he goes on to ask why Kantians and consequentialists nevertheless disagree, and I think his answer is exactly on the right track:

Why do Kantians and consequentialists then disagree despite this? Perhaps the best way to understand why they still disagree is to think that they have different views about what makes the intuitively right acts right. Consequentialists claim that the acts which we all believe to be right are right because they bring about the best outcomes. In contrast, Kantians claim that these acts are right because the relevant maxims for them can be willed to be universal laws. Consequentialists and Kantians give competing explanations for why certain acts are right even if they can agree on which acts are right.²²

20 Parfit, *On What Matters*, 1:419.

21 Indeed, the manuscript that was widely circulated before the publication of *OWM* was titled "Climbing the Mountain."

22 Suikkanen, *This Is Ethics*, 104.

What Suikkanen is bringing to our attention here is that moral theories are not only in the business of producing the correct particular deontic verdicts. They also seek to explain *why* certain acts are right or wrong (obligatory, forbidden, or allowed). This attempt at explanation I take to be at the very heart of moral theorizing. Indeed, ethicists often seem to be more interested in getting the explanation right than in fine-tuning their theory's output.

What kind of explanation does Suikkanen refer to here? Certainly not the *metaethical* one, having occupied philosophers since Harman, of whether purported moral facts figure in our best explanations of our moral beliefs.²³ Instead, it is a more generic notion of explanation inherent to normative ethics, pertaining to the reasons we are giving within ethical discourse and practice. Different ways of understanding this kind of explanation present themselves and we have to be careful here since not all accounts fit well within Parfit's framework. For example, Berker has recently proposed that moral explanations are quite naturally amenable to an analysis in terms of the notion of *grounding*.²⁴ On such an analysis, the alternative traditions of moral theorizing make claims as to what is fundamentally prior in ethics. Importantly, that priority is neither epistemic nor causal or conceptual.²⁵ Instead, according to that view, moral explanations identify the ultimate metaphysical grounds for an act's rightness or wrongness.

The problem with such an analysis is that Parfit, as we have seen, defends a non-metaphysical version of moral realism.²⁶ Presumably, this also includes explanatory claims. In order to stay within Parfit's framework, it is thus better to adopt a non-metaphysical understanding of moral explanation. I suggest that the most straightforward is a semantical one. Seen this way, moral explanations consist of truth-apt sentences (or propositions) about why acts are right or wrong. This is in accordance with Parfit's remarks on other moral claims. Despite having no ontological implications, Parfit informs us, such claims can nevertheless be true in the strongest sense.²⁷ Thus, just as we do not make any ontological commitments when attributing to acts the property of rightness, saying that some feature makes an act right or wrong does not entail any ontological consequences.

How do these different explanations relate to one another? We can begin by observing that they use markedly different language. Kantians typically refer to such concepts as autonomy, good will, humanity as an end in itself, etc. Con-

23 Harman, *The Nature of Morality*.

24 Berker, "The Unity of Grounding."

25 Compare Berker, "The Unity of Grounding," 10–13.

26 Compare Parfit, *On What Matters*, 2:486.

27 Parfit, *On What Matters*, 2:479.

tractualists use the ideas of a (hypothetical) agreement, some kind of initial position, and cooperation between rational (self-interested) agents, among others. Finally, the notions that consequentialists emphasize include that of an (actual or foreseeable) outcome, its overall utility, and its ranking on a scale of goodness. Most ethicists think that the differences in these explanations reach deeper than the terminological surface. They agree with Suikkanen that the different traditions give *competing* explanations.

This might be mistaken, though. There is a possibility that the different traditions are merely *notational variants*. When putting forward seemingly incompatible claims about what makes acts right, Kantians, consequentialists, and contractualists would merely be using different terminology. A better understanding of these claims might indicate a way of translating one explanation into the terms of the other and demonstrate that there is no real incompatibility after all. These matters are not to be conclusively decided here. Since a translation of predicates could be rather complicated and surprising, we cannot dismiss that possibility out of hand. However, I contend that this is not what is called for at this stage. First, it is important to note that the notational variants view clearly goes against the orthodox picture. There is no obvious way to translate the explanatory notions used by one of these frameworks into those used by its rivals. Standard textbooks conceive of the different traditions as rival, incompatible frameworks.²⁸ The onus is therefore on the critic of the standard view to show that the three traditions are in fact not giving competing explanations, despite it obviously seeming like they do. Parfit certainly has not shown this.

Second, there is good additional reason to think that it could not be shown. It seems that it is constitutive of the moral traditions that they give explanations of a certain kind. A consequentialist theory that does not maintain that the rightness of an act is an upshot of its outcomes alone is (obviously) not a consequentialist theory. What is more, a Kantian theory that would accept this *same claim* about the rightness of an act is not a Kantian theory. Thus, there is a very strong sense that these theories give mutually exclusive explanations *and* could not cease to do so without losing a constitutive feature. This finds its expression in the fact that the different traditions are sometimes even *defined* in opposition to each other.²⁹ Pending arguments to the contrary, we thus have very strong

28 For recent examples, compare Tännsjö, *Understanding Ethics*; Driver, *Ethics*; and Deigh, *An Introduction to Ethics*.

29 For instance, in "Deontological Ethics," Alexander and Moore use consequentialism as a foil to define deontology.

reason to consider these alternative theories to be logically incompatible when it comes to their explanations.³⁰

If this is correct, Parfit is not justified in claiming that all differences among the alternative traditions have been resolved. Though Kantians, contractualists, and consequentialists might agree on a set of principles (and thus on the deontic verdicts that follow from them), they put forward different, mutually exclusive explanations for why those are the correct principles. This leads us to a puzzling result. On the one hand, alternative traditions of moral theorizing seem to be able to lead to the same deontic verdicts. On the other hand, differences remain as to how those theories explain the deontic status of acts. In other words, those theories are *deontically equivalent while at the same time being logically incompatible*. How are we able to explain this?

3. UNDERDETERMINATION AND REALISM

Remarkably, in the philosophy of science a similar phenomenon has attracted a lot of attention for quite some time now: the underdetermination of theory by evidence. This analogy between science and ethics has recently been pointed out by Dietrich and List.³¹ Since it takes a prominent role in our argument, it will be useful to provide some background to the phenomenon in science before considering the analogy.

3.1. Underdetermination in Science

Following the work of Duhem and Quine, philosophers of science have debated a highly influential, though nonetheless controversial, idea: *the underdetermination of theory by evidence*. Duhem and Quine both held that sometimes alternative scientific theories are able to accommodate exactly the same observable evidence, while simultaneously making radically different assertions about unobservables.³² These theories thus give mutually exclusive explanations for the

30 Since this is quite a mouthful, I will also speak of *theoretical* or, more specifically, *explanatory incompatibility* when I want to highlight that the incompatibility is not regarding the deontic content. This does not designate a special kind of theoretical/explanatory incompatibility but is rather a short form for logical incompatibility concerning the theoretical/explanatory claims.

31 Dietrich and List, "What Matters and How It Matters."

32 Duhem's *The Aim and Structure of Physical Theory* (*La Théorie Physique: Son Objet et sa Structure*) is widely regarded as the *locus classicus* for the underdetermination thesis. Quine's "Two Dogmas of Empiricism" introduced the idea to the analytic tradition, while his own views have kept changing over time. (Compare Quine, "Two Dogmas of Empiricism," "On

same evidence. They are, in other words, *empirically equivalent while at the same time being logically incompatible*.³³

The standard examples to illustrate underdetermination are mostly drawn from the history of science, especially physics. Famously, at the time of Copernicus, the choice between his new theory and its Ptolemaic alternative was underdetermined by observational data. Likewise, not that long ago, the relevant observations seemed to be compatible with both a corpuscular and a wave theory of light.³⁴ For our purposes, it would take us too far off track to consider them more closely. However, the basic idea can be illustrated by means of everyday phenomena. Ladyman invites us to consider the case of some person at a train station waiting for a delayed train.³⁵ While waiting, this person develops several hypotheses on why the train might be delayed, such as problems with the engine, a staff shortage, or a signal failure. All the hypotheses are compatible with the data available to the person at that time, but the explanations are incompatible (barring the case of multiple causation). Of course, such a case of underdetermination is not very interesting for the obvious reason that it is a consequence of the epistemic situation of a specific person at a specific time. Other people probably already know the correct explanation and the person herself could find it out easily with some research. What makes for a philosophically more interesting case are situations in which the whole scientific community, given all the data available, cannot decide among the different theories. This, or so it is claimed by proponents of underdetermination, has been (and still is) the case with at least some of our scientific theories.

Not much more is common ground besides this. Duhem and Quine already held quite different views about the nature and scope of underdetermination. Whereas Duhem argued for a moderate, restricted thesis that was informed by detailed historical examples, Quine envisioned a more radical version that

the Reasons for Indeterminacy of Translation,” “On Empirically Equivalent Systems of the World,” and “Three Indeterminacies.”)

- 33 At a later stage, Quine renounces his claim of *logical* incompatibility between the rival theories (“On Empirically Equivalent Systems of the World”). Instead, he now holds that the incompatibility between the empirically equivalent theories merely consists of us not being able to find a way to reconcile them by way of reconstruing their predicates. That is, the theories are not logically incompatible; they are simply using different predicates that we have not yet managed to translate into each other. For reasons I explain in note 42, anti-realists should resist this move and insist on there being logical incompatibility.
- 34 Both these cases have already served as examples of underdetermination to Duhem. More recently, Stanford, *Exceeding Our Grasp*, and Bonk, *Underdetermination*, have discussed a range of further cases at length.
- 35 See Ladyman, *Understanding Philosophy of Science*, 162–63.

pertains to the totality of our knowledge, and for whose justification he relied more heavily on general epistemological, logical, and linguistic considerations.³⁶ Many authors have adopted Quine's wider understanding, so that the thesis that is most often discussed today looks something like this:

Underdetermination of Scientific Theory (UST): There are alternatives to even our best scientific theories that can account for exactly the same evidence while containing incompatible propositions.³⁷

As might be expected, considering how long philosophers have grappled with the idea of underdetermination, a fair amount of criticism has also been directed at it. These criticisms can be roughly divided into two classes. On the one hand, philosophers have questioned whether underdetermination is in fact a sufficiently widespread phenomenon to justify a thesis like UST. Thus, it has been pointed out that, for most of the time, the discussion has focused on a very small set of examples, which are mostly from a subset of physics, and which are not representative of science *per se*.³⁸ This kind of criticism, though of great relevance for the debate in science, should not bother us too much, and for a simple reason. As we shall shortly see, I am not advancing a claim to the effect that *because* there is underdetermination in science we should also expect to find it in ethics.

On the other hand, there are also more general lines of criticism that cast doubt on the mere possibility of rendering the idea of underdetermination plausible (or coherent) at all.³⁹ Such criticism is more dangerous to our purpose because it might generalize to any form of underdetermination in any area of inquiry. I address some of these objections below. For now, I will simply continue on the assumption that underdetermination can be rendered a coherent idea.

36 For a similar assessment of what the differences between Duhem and Quine amount to, see Pietsch, "Defending Underdetermination or Why the Historical Perspective Makes a Difference."

37 A very concise overview of the different versions that have been proposed can be found in Park, "Philosophical Responses to Underdetermination in Science."

38 See Kitcher, *Real Realism*, and Stanford, *Exceeding Our Grasp*.

39 These include the claims that underdetermination is a nonstarter because the distinction between observables and unobservables cannot be upheld (Maxwell, "The Ontological Status of Theoretical Entities"), that most versions of the underdetermination thesis are based on a wrong-headed equation of the logical consequences of a theory with its evidential support (Boyd, "Realism, Underdetermination, and a Causal Theory of Evidence"; Laudan, "Demystifying Underdetermination"), and that the notion of empirical equivalence cannot be rendered precise in a way that would support the underdetermination thesis (Laudan and Leplin, "Empirical Equivalence and Underdetermination," and Worrall, "Underdetermination, Realism and Empirical Equivalence").

3.2. *The Anti-Realist Argument from Underdetermination*

On that note, let us turn to the more pressing issue for our purpose, which is the upshot of underdetermination for the realism debate. Why is underdetermination considered to be dangerous for realists? At first glance, it merely poses a problem for theory choice. Although we know that we are facing rival theories, since they contain incompatible propositions, our choice between them is rendered indeterminate by the fact that they can account for exactly the same evidence. Yet, in addition, many philosophers have also held that underdetermination poses a specific problem for scientific realism.⁴⁰

To get a firm grasp of what the danger to scientific realism amounts to, we can start with a thin definition of scientific realism modeled on Sayre-McCord's definition of that doctrine for the moral realm:

- a. Scientific claims, when literally construed, have truth values.
- b. At least some scientific claims actually are true.

Note again that, according to such a thin definition, there is neither a metaphysical nor an epistemological component to realism. However, as is the case in ethics, most scientific realists will add such additional components to their preferred view.⁴¹ It is important to see that whether underdetermination poses a threat to a realist view depends on which component that is and how strong it is formulated.

Before I can go on spelling that out, we first need a formulation of the basic idea behind the anti-realist argument on the table. So as not to complicate matters too much, I will outline an intuitive version of the argument and postpone discussion of its complications to a later stage. Hence, the argument could read something like this:

- P1. If two scientific theories (ST) can account for exactly the same evidence, it is equally reasonable to believe either of them.
- P2. If it is equally reasonable to believe either of two ST, we have no reason to attribute truth to one but not the other.

40 Indeed, the argument from underdetermination is sometimes seen as one of two main arguments against scientific realism, together with the so-called *pessimistic metainduction*. Compare Bortolotti, *An Introduction to the Philosophy of Science*, and Stanford, "Underdetermination of Scientific Theory."

41 Compare Newton-Smith and Lukes, "The Underdetermination of Theory by Data," and Bortolotti, *An Introduction to the Philosophy of Science*. Most scientific realists would probably not even see the point in defending the doctrine if described so thinly.

- P3. If two ST contain incompatible propositions, they cannot both be true.⁴²
- P4. If two ST cannot both be true, and we have no reason to attribute truth to one but not the other, then none of them should be considered true.
- UST. There are alternatives to even our best scientific theories that can account for exactly the same evidence while containing incompatible propositions.
- C. Therefore, even our best scientific theories should not be considered true.

Most readers will probably have immediate reservations about one or more of the premises. I must ask them to bear with me for a little longer. Here, I only want to draw attention to how the argument relates to the different versions of realism that we identified in the beginning.

First, note that the argument, as it stands, does not directly challenge scientific realism as we defined it. On our definition, scientific realism is a position about scientific claims, not theories. However, we can reasonably argue that, since our scientific claims follow from our best theories and the latter should not be considered true, we also should not believe the former. Still, there is a further problem. The argument is evidently of the *skeptical* variety due to its focus on the reasonableness of belief. It states that we should not consider the theories (or their claims) true because, based on the evidence, we have no reason to prefer one theory and we also know that they cannot both be true. But insofar as this poses a problem for realism, it presupposes that there is an epistemic component to that position.⁴³ Moreover, the strength of the argument is inversely proportional to the strength the realist ascribes to the epistemic component. The stronger the realist insists on us being able to find out about scientific truths, the

42 Here is the reason why anti-realists need to insist on logical incompatibility between the rival theories. If incompatibility comes down to nothing more than our practical inability to reconstrue predicates, we would have no reason to believe P3. That being said, the burden is not automatically on the anti-realist to prove logical incompatibility. If there is a strong initial indication that there is logical incompatibility between the theories, it is on the opponent of the argument to show that the seeming incompatibility can be resolved.

43 But recall note 12. The anti-realist can try and strengthen the argument by charging that the lack of any ability to attain knowledge about moral truths renders baseless the idea that there are such truths. The argument, in combination with this claim, would thus also pertain to forms of realism that include only a metaphysical component. This surely needs to be taken into consideration when assessing the strength of the argument beyond our present focus on Parfit.

more damning the argument will prove if it indeed succeeds in establishing that we cannot find out about those truths. This will shortly become very important.

Second, note that the argument, as it stands, is deliberately formulated in terms of the truth of propositions and theories. It does not depend on any metaphysical assumptions. This gets overlooked easily because philosophers of science are naturally tempted to express the idea of underdetermination in terms of unobservable objects, thus in ontological terms. But it is commonly acknowledged that the underlying problem is a broader epistemological one.⁴⁴ We can therefore resist this tendency and formulate the challenge solely in terms of the truth of (scientific) theories and propositions. This opens up the possibility of transferring it to other domains. More specifically, it opens up the possibility of transferring it to debates about forms of realism that are ontologically noncommittal.

3.3. *Adapting the Argument to the Realm of Ethics*

If we are to believe anti-realists in the scientific domain, empirically equivalent theories that contain mutually exclusive propositions pose a problem for scientific realism. But how does this relate to our case in ethics? This is where Dietrich and List come into play.⁴⁵ Drawing on decision-theoretic work, they propose a new formal framework for the classification of moral theories. In that context, they introduce what they call the *reason-based representation* of moral theories. Echoing Suikkanen, this framework distinguishes between two dimensions of moral theories:

Reason-based representations encode not only a theory's *action-guiding recommendations* (that is, how we should act, according to the theory), but also the *reasons* behind those recommendations (that is, why we should act in that way).⁴⁶

Dietrich and List's primary use of their framework is for a formal taxonomy of moral theories, which need not occupy us here. However, it also helps to drive home an important analogy. As they point out themselves, their framework "shed[s] light on an important but still underappreciated phenomenon . . . : the *underdetermination of moral theory by deontic content*."⁴⁷ In their view, just as scientific theories can be underdetermined by the empirical evidence, moral

44 See Ladyman, *Understanding Philosophy of Science*; Stanford, "Underdetermination of Scientific Theory."

45 Dietrich and List, "What Matters and How It Matters."

46 Dietrich and List, "What Matters and How It Matters," 422.

47 Dietrich and List, "What Matters and How It Matters," 422.

theories can be underdetermined by their deontic content. By formally distinguishing between the two dimensions, they show that it is at least theoretically possible that theories would differ when it comes to the second dimension but not the first. Moral theories can give different explanations of what makes acts right or wrong but nevertheless agree about the deontic status of those acts. This is, in a structural way, similar to what has been discussed in the underdetermination debate in the philosophy of science.

The analogy relies on treating both the data of scientific theories as well as the deontic verdicts of our moral theories as *extensions* of those theories. Dietrich and List do not flesh out in detail what this amounts to. However, it seems clear that it does not mean that the particular verdicts are epistemically on a par with the data of scientific theories *in every respect*. First, it is not being claimed that *particular* verdicts are prior in the sense that they have an initially higher credibility than, e.g., mid- or highest-level principles. What the analogy presupposes is only that particular verdicts have to be accounted for by moral theories, just as the data has to be accounted for by scientific theories. Second, it is not being claimed that particular *verdicts* are the ultimate ground of epistemic justification in ethics. Instead, what is proposed here is compatible with the view that our moral verdicts are themselves grounded in, for example, (non-doxastic) intuitions of the same content.⁴⁸

More importantly for our present case, by ascribing particular verdicts the role of evidence in ethics, we do not presuppose anything that Parfit does not already acknowledge. We will shortly contemplate whether particular verdicts are the *only* evidence in ethics. But unless we deny that the particular moral verdicts play any role whatsoever in theory choice in ethics, the analogy stands. Parfit definitely does not deny this, since otherwise he would not have to worry about alternative theories arriving at different deontic verdicts in the first place.⁴⁹

Though Dietrich and List's main focus is on the so-called *consequentializing* debate, they do comment on Parfit's project in *OWM*. They contend that their rea-

48 Indeed, talk of the *data* of moral theories has always had particularly wide currency in intuitionist theorizing. Compare Ross, *The Right and the Good*, 41, for a classical statement of that view, and Audi, "Intuition, Inference, and Rational Disagreement in Ethics," 476, for a more recent one.

49 One might object to the analogy on a different ground. Perhaps one does not think that moral theories are supposed to account for the moral verdicts that we *actually*, at this time, hold. Instead, they tell us what verdicts we *should* be holding. But this objection misses the point. My aim is to show that Parfit's solution gets him into problems. For this purpose, it is sufficient to point out that Parfit gives the verdicts that we do actually hold a central role in deciding which theories are correct. Indeed, much of volume one of *OWM* is a constant refinement of the different theories in order to make them compatible with specific cases.

son-based representation supports at least the possibility of different theorists climbing the same mountain on different sides, while remaining uncommitted on whether this is indeed the case with Parfit's preferred theories.⁵⁰ However, they remain largely silent when it comes to the metaethical consequences of this observation.⁵¹ Yet, if the proposed analogy stands, it is easy to see how we can construct a structurally analogous anti-realist argument for the moral realm as was conceived for the scientific realm. We only have to substitute *moral theories* for *scientific theories* in the argument above to see this:

- P1'. If two moral theories (MT) can account for exactly the same evidence, it is equally reasonable to believe either of them.
- P2'. If it is equally reasonable to believe either of two MT, we have no reason to attribute truth to one but not the other.
- P3'. If two MT contain incompatible propositions, they cannot both be true.
- P4'. If two MT cannot both be true, and we have no reason to attribute truth to one but not the other, then none of them should be considered true.
- UMT. There are alternatives to even our best moral theories that can account for exactly the same evidence while containing incompatible propositions.
- C. Therefore, even our best moral theories should not be considered true.

Ethics might thus face a similar problem as science. We find theories with the same extension that nevertheless contain propositions that cannot be true at the same time. If we are unable to choose between those theories, that makes it difficult for us to believe the claims they put forward. Instead of helping realism by refuting the argument from disagreement, Parfit might have laid the ground for a new anti-realist challenge.

4. THREE POSSIBLE REALIST REJOINDERS

So far, I have only presented a very basic version of the anti-realist argument and made a suggestion as to how to transfer it to the moral realm. To add some more depth to this, I am going to discuss three possible realist rejoinders, which take

⁵⁰ See Dietrich and List, "What Matters and How It Matters," 451.

⁵¹ They mention how a parallel view to scientific instrumentalism in ethics might clash with their reason-based representation (Dietrich and List, "What Matters and How It Matters," 425–26). However, they do not pursue this line of thinking any further.

their inspiration from similar objections in the scientific debate. They bear on P₁, P₂, and P₃, respectively.⁵² Even though I will ultimately argue that Parfit is not in a position to make use of any of them, it should also become clear that the situation is much more complicated than I have been able to convey so far. Thinking in terms of underdetermination is apt to raise some intricate new issues.

4.1. Additional Evidence and Theoretical Virtues

P₁ invites two kinds of criticism. First, according to the proposed analogy, it is the theories' *deontic* verdicts alone that take the place of empirical evidence. However, it might sensibly be objected, moral theories also generate non-deontic verdicts, such as *axiological* ones. For example, they might also yield verdicts on the goodness and badness of acts, of states of affairs, or of the characters of agents. Yet if one theory were to yield much more plausible verdicts of such a kind than any of its rivals, these might well tip the balance in its favor. The overall argument would thus fail because it relies on too restricted an account of what constitutes the evidence in ethics.

There is one swift reply available to this objection if we focus on the present context only. Parfit thinks that his convergence argument helps realism by settling the deontic quarrels. However, if differences remain about other verdicts, the proposed convergence is much less effective in countering the argument from disagreement. Moreover, if the rest of our observations are correct, the theories are incompatible and we still need to decide among them. Even if there remain axiological judgments to decide the case, we can no longer do so on the basis of deontic ones. There is, in sum, *less* of a basis to distinguish between the theories if the convergence argument proves successful. Settling the deontic conflicts thus makes it *more* difficult to adjudicate between the incompatible theories and thereby more difficult to claim that we can know some truths in ethics. Hence Parfit's convergence argument does not advance the case for realism.

However, if we look beyond Parfit and ask whether options remain for other realists to counter the argument from underdetermination, the objection gains more traction. Perhaps the correct deontic verdicts can indeed be accounted for by incompatible theories, yet that still leaves open the possibility of choosing among them on the basis of (a class of) other verdicts they yield. Since the aim of this paper is not to defend a general version of the underdetermination argument, I will not be able to counter this objection in detail. However, I want to

⁵² I will not consider objections to P₄, which I take to be the least controversial. I am also not claiming that these are the only possible objections, though I do think that they are among the most important ones.

at least hint at two possible routes that the anti-realist's reply can take. One is to insist that deontic verdicts or intuitions enjoy a privileged status. This might be the case because our intuitions about them are more firm, conferring a lesser status to our intuitions about, e.g., axiology. Anti-realists thus needed to defend a moral epistemology that vindicates the primacy of deontic verdicts. Alternatively, it might have something to do with the purpose of moral theories themselves. If moral theories are, in an important sense, *practical*, deontic verdicts will have a privileged status because action-guiding verdicts follow from them, not from axiological ones.

Another route is to agree in principle that other evidence can play a role in theory choice, while maintaining that it is insufficient to actually tip the balance in the present case. The anti-realist would thus have to show that, if there are differences regarding axiological verdicts, none of those speak decisively in favor of one of the competing theories. Maybe the rival theories each get equally important subsets of those other verdicts correct. Bringing in additional evidence does not guarantee that we will be able to choose.

The second kind of criticism takes issue not so much with what we put in place of the empirical evidence, but with the idea that the *empirical evidence alone* can adjudicate among the rival theories. We have seen that the anti-realist tries to exploit the fact that different theories can account for exactly the same evidence, to argue that it is equally reasonable to accept any of them. But, so it has been argued in the scientific case, this is misleading, at least if we consider the evidence to narrowly consist of the empirical data. Many critics have objected to what they see as an impoverished understanding of scientific methodology underlying the anti-realist argument, to the effect that it identifies the deductively deducible consequences of a theory with its evidential support. To put it in a catchphrase, they claim that empirical equivalence does not entail evidential equivalence.⁵³ Instead, theories exhibit additional *theoretical virtues* by which we compare them, for example simplicity, predictive fruitfulness, or non-*ad-hoc*-ness. Such theoretical virtues might well be brought in to decide between theories that are empirically equivalent.

Anti-realists have reacted to this objection in basically two ways. One way is to deny that theoretical virtues are relevant at all to the truth of a theory. Such virtues, it is argued, give us a pragmatic criterion for which theory to use, but whether a theory is, e.g., simple does not have anything to do with it being true.⁵⁴

53 Critics of underdetermination that have contributed to driving home this point include Boyd, "Realism, Underdetermination, and a Causal Theory of Evidence"; and Laudan, "Demystifying Underdetermination."

54 Van Fraassen has very famously defended this line of argument (*The Scientific Image*).

The second way for an anti-realist to react is by pointing out difficulties for the project of deciding between theories on the basis of such additional criteria. Tulodziecki argues that we might face very difficult problems of weighing different virtues when they are exhibited by rival theories, even problems of incommensurability.⁵⁵ The debate about these issues remains open. Yet, at least in principle, it is easy to see how such considerations could be brought into play in the moral case as well. So far I have relatively uncritically assumed that deontic (and possibly axiological) equivalence suffices to get the anti-realist argument flying. However, moral theories exhibit theoretical virtues as well, and those might tip the balance in favor of one of the rival traditions.⁵⁶ I am not going to take sides in this dispute. For what it is worth, I do think that, no matter whether theoretical virtues are indeed relevant to the question of truth, further investigation into them would be of value in ethics. They have not received as much attention yet as they deserve. In ethics, we generally take for granted that the different theoretical traditions arrive at very different conclusions about deontic verdicts. This renders it unnecessary to search for further grounds on which to decide among them. One especially interesting question that Parfit's convergence argument opens up is whether theoretical virtues might not break the tie among the rival traditions.⁵⁷

However, if we narrow down the question again to whether there is a strategy that suits Parfit specifically, I am fairly sure that this is not it. The reason for that is that it would jeopardize what could be called his *reconciliatory project*. Parfit is explicit that rather than proposing a new moral theory, he wants to learn from existing ones.⁵⁸ His is not a reductionist project but one of reconciliation among the major moral traditions. Hence, all three mountaineers reach the summit. But by referring to additional theoretical virtues we would, if successful, decide which of the deontically equivalent theories is the correct one after all.⁵⁹ The whole point of this strategy is to reach a decision *for one* of the rival theories. If theoretical verdicts could decide the case, we would no longer have reason to believe in the three traditions of moral theories, but only in one of them. The price that the discussed strategy takes for saving Parfit's project from the argument

55 Tulodziecki, "Epistemic Equivalence and Epistemic Incapacitation."

56 Hooker offers an influential line of reasoning for rule consequentialism along these lines (*Ideal Code, Real World*).

57 Carrier suggests that underdetermination in science fulfills a similar function, in that it can serve as a test-tube to lay open the nonempirical virtues that play a role in theory choice ("Underdetermination as an Epistemological Test Tube").

58 Compare Parfit, *On What Matters*, 1:174.

59 The same point holds for bringing in additional axiological evidence.

from underdetermination would thus be to give up on reconciliation. Since I take the reconciliatory project to be very dear to Parfit's heart, I do not think that he would want to opt for this strategy.

This points to a conflict at the heart of Parfit's project. Parfit wants to have his cake and eat it, too. He wants to get rid of the disagreements that threaten realism, but he does not want to make up his mind about which moral theory is the correct one. Thus he undertakes much effort to clear away any deontic disagreements, making it more difficult in the process to choose among the theories. After the convergence argument has gone through, it is harder (if possible at all) to find the correct theory for lack of a deontic basis on which to decide. However, this does not seem to bother Parfit, which is baffling. Indeed, if we take heed of the lessons from the philosophy of science, it looks like we end up with a picture that is more congenial to anti-realist views. Faced with mutually exclusive theories that are extensionally equivalent, the typical realist reaction is to look for other criteria to decide the case. It is anti-realists who have pointed out problems for this proposal and who are generally much more comfortable with different, noncompatible explanatory frameworks. Thus, Parfit's whole project of reacting to disagreements by showing that different theories can account equally well for them has a distinctly anti-realist ring to it.

4.2. *What We Know and What Is True*

Except, what if we do not need to bother about being able to find the correct theory? The second objection goes right at the epistemic component we identified as one of the possible additions to the thin version of realism. P2 draws a direct connection between the reasonability of some of our beliefs about a subject matter and our attribution of truth to that same subject matter. This, a realist might reply, seems an overly hasty inference. There is a perfectly acceptable alternative. We could simply claim that we have not, or will never have, the evidence to adjudicate among some theories, but that we are nevertheless justified in believing that only one of them is true. That is, we could hold on to the conviction that there are facts, and correspondingly possible reasons, inaccessible to us, that decide which theory is true.⁶⁰ Would Parfit want to opt for a similar strategy in the moral case? That is, could he hold that, in the case of underdetermination, we are ignorant about which of the theories is correct, while at the same time insisting that one of them is?

60 There is much literature on this in the philosophy of science. A highly illuminating discussion is between Newton-Smith and Lukes ("The Underdetermination of Theory by Data") and Bergström ("Underdetermination and Realism") on the so-called *ignorance response* to underdetermination.

I have already pointed out that the fact that Parfit takes disagreements as seriously as he does is a strong indication that he subscribes to a strong epistemic understanding of realism, which suggests that he would not want to argue along these lines.⁶¹ But we need not turn to such indirect clues. As it turns out, such an answer is not compatible with Parfit's explicitly stated views on moral epistemology. Witness the following passage:

If we had strong reason to believe that, even in ideal conditions, we and others would have deeply conflicting normative beliefs, it would be hard to defend the view that we have the intuitive ability to recognize some normative truths. We would have to believe that, when we disagree with others, it is only we who can recognize such truths. But if many other people, even in ideal conditions, could not recognize such truths, we could not rationally believe that we have this ability. How could *we* be so special? And if none of us could recognize such normative truths, we could not rationally believe that there *are* any such truths.⁶²

As these remarks show, it would not be in the spirit of Parfit's moral epistemology to opt for a solution along these lines. Importantly, it is not Parfit's commitment to the claim that we have intuitive abilities to recognize some normative truths that renders his position vulnerable to the skeptical attack. Intuitionists can consistently hold on to the idea that we have an intuitive ability to recognize truths while admitting that this does not guarantee that we will be able to agree about them. Rather, it is Parfit's conviction that if it turns out that we do in fact disagree substantially about some of our deep moral beliefs, we are no longer

61 Whether this is itself a reasonable conviction is an open question. Many commentators have stressed that Parfit's fear of disagreements, which seems to underlie those epistemological claims, might be overstated. Several have also pointed out that there is a discrepancy between Parfit's heightened uneasiness with disagreements about normative concerns and his more confident reaction to metaethical disagreements. See Larmore, "Morals and Metaphysics"; Darwall, "Agreement Matters"; and Smith, review of Derek Parfit, *On What Matters*, vol. 2.

62 Parfit, *On What Matters*, 2:546. The passage might seem out of place, since Parfit is here talking about our ability to recognize normative truths, whereas my objection is premised on him being committed to our ability to recognize explanatory truths (what makes acts right). I will attend to this worry shortly. However, there are other passages where Parfit makes similar suggestions that are not restricted to normative disagreements, such as the following: "Such disagreements give us reasons to doubt that we are the people whose beliefs are true. These disagreements may also give us reasons to doubt that any of the conflicting beliefs are true. Perhaps none of us is right, because our questions have no answers" (2:427–28). In this passage, Parfit is referring to disagreements about "what it would *be* for things to matter, and about whether anything *could* matter" (2:427). Thus, it does not seem to me that my proposed reading is an unfair one.

entitled to trust in these abilities. Since on his view there are no other ways to recognize moral truths, Parfit is convinced that we would then no longer have reason to believe in such truths at all.

Of course, Parfit is not committed to the view that we must already have found out about those truths, since he adds the proviso that we only have to be able to do so *in ideal conditions*. But this does not change the picture. Parfit acknowledges that the different traditions come to the same set of deontic verdicts; indeed, that is the whole point of his convergence argument. Since the deontic consequences of those traditions are thus identical, there could be no way, even in ideal conditions, of distinguishing among them on that basis. Moreover, if my reply to the first objection is correct, Parfit would not want to appeal to any other criteria, either. Yet there are nevertheless deep disagreements between the alternative traditions when it comes to their explanatory content. We thus have strong reasons to believe that we and others would, even in ideal conditions, disagree *at least insofar* as we cannot adjudicate among our theories' incompatible explanatory claims.

4.3. *Do the Explanatory Disagreements Matter?*

This leads us to the final objection. We have seen that the problem for the realist arises because, although it might be possible to put the disagreements about our *deontic* principles (and with that about all specific cases) to rest, other disagreements persist regarding explanation. P₃ states that two theories cannot both be true if such disagreements persist. However, to the realist, maybe this is not what we were looking for in the first place. What if we were simply to claim that we are only concerned about the deontic disagreements and not the explanatory disagreements? That is, what if we were to claim that our position only entails that there are truths about our deontic principles and verdicts, but not about why these principles and verdicts are the correct ones? Parfit might in such a spirit restrict the importance of our coming to find moral truths to the deontic realm. Since there are no remaining differences there, his line of reasoning would then be unaffected by the anti-realist argument.

The textual basis in *OWM* on this issue is rather thin. At least one passage strongly suggests that Parfit might consider such a solution to the problem. After outlining in great detail how distorting influences are responsible for many of our disagreements about deontic verdicts, he makes the following claims:

Some other moral disagreements are not about *which* acts are wrong, but about *why* these acts are wrong, or what *makes* them wrong. Different answers are given by different systematic theories, such as those developed

by Kantians, Contractualists, and Consequentialists. Such disagreements do not directly challenge the view that we are able to recognize some moral truths. In defending this view, it is enough to defend the claim that, in ideal conditions, there would be sufficient agreement about which acts are wrong. Though we also have intuitive beliefs about why many acts are wrong, and about the plausibility of different systematic theories, we would expect there to be more disagreement about these other questions. As I have also argued, however, when the most plausible systematic theories are developed further, as they need to be, these theories cease to conflict. If that is true, these theoretical wars would end.⁶³

Parfit does not explain why we would expect there to be more disagreements about explanatory questions. Also, if I am correct, he has not shown that the theories would cease to conflict. Nevertheless, the passage suggests that he does consider some kinds of disagreements, such as the ones about explanation, not to be of equal importance to the realist's case.

I am not sure how much importance to accord to this passage. Yet even if this turned out to be Parfit's preferred solution, we need to ask whether it is a tenable one. Two reasons speak strongly against it. First, the solution seems *ad hoc*. Parfit is of course right that the explanatory disagreements do not logically contradict the position that there are some moral truths. However, at least at first sight, the explanations we ordinarily put forward for why certain acts are right or wrong seem to be meant to be taken at face value. This generalizes to our theories. Typically, our moral theories are not only trying to enumerate and systematize our particular verdicts, but also have an explanatory aspiration. Barring further arguments to the contrary, we should take these explanations at face value, too. At the very least, the burden of proof is on the realist to prove otherwise.

The impression of *ad-hoc*-ness is reinforced by the fact that Parfit does consider some disagreements other than deontic ones to be threatening to realists:

Disagreements are deepest when we are considering, not the wrongness of particular acts, but the nature of morality and moral reasoning and what is implied by different views about these questions. If we and others hold conflicting views, and we have no reason to believe that *we* are the people who are more likely to be right, that should at least make us doubt our view. It may also give us reasons to doubt that any of us could be right.⁶⁴

63 Parfit, *On What Matters*, 2:554

64 Parfit, *On What Matters*, 1:418–19.

Thus, according to this passage, there are also deep disagreements about other aspects of morality and those seem to give rise to the same worries as the ones about deontic verdicts. The fact that Parfit is talking about the nature of morality and moral reasoning here suggests that he is thinking of metaethical differences, not of the explanative ones that I have in mind. But the general point stands nevertheless. If such disagreements are indeed a threat, what makes our explanatory disagreements so special that they are not? Why should we not be bothered by the fact that we cannot find out which explanation of the deontic status of our acts is correct?⁶⁵

Comparison with the scientific case hints at an additional, deeper reason. There is a corresponding discussion in the philosophy of science. It has been asked whether one can be a realist about the observable and remain agonistic about unobservables. Yet, strikingly, this move is generally taken by *anti*-realists. Thus Van Fraassen, who has famously argued that science does not aim for more than empirically adequate theories, is also very explicit about his own constructivist empiricism being an alternative to scientific realism.⁶⁶ What he denies is that the entities our theoretical concepts refer to in order to explain the data have the same entitlement to be taken at face value. Most scientific realists accept this challenge. They seem to think that a realist position that deserves its name cannot restrict its realism to the claims that theories make about what we can all readily observe, but also has to apply to the more theoretical claims. The pressing realist questions are in the end not about the data itself, but about the further claims that we make to account for them.

Does this generalize to ethics? I am not sure. There is no obvious reason why the criteria for realism should be the same over all domains of inquiry. Instead, realism might come in domain-specific variations. Maybe in ethics we can be satisfied with knowing which acts are right or wrong without knowing why this is so. But again, the realist would at least have to give us an explanation for the asymmetry with the scientific case. Why is the glass taken to be half full in ethics, whereas it would in similar cases be considered half empty in science? Barring further arguments, the idea of restricting the aim of our moral theories to deontic adequacy looks suspiciously similar to the one of restricting the aim of

65 Note again that our definition of realism does not commit a realist to hold that there is a fact of the matter to every moral question. However, to exclude a whole class of statements (those about moral explanation) from being truth-apt seems arbitrary. Usually, realists argue that when we are not able to find out the truth about some moral question this is explainable by citing phenomena such as vagueness. It would yet have to be shown why explanatory statements summarily suffer from such an impairment.

66 Compare Van Fraassen, *The Scientific Image*.

our scientific theories to empirical adequacy, which is, after all, an anti-realist suggestion.⁶⁷

5. CONCLUDING REMARKS

Let me finish by taking stock and tying up some loose ends. Rather than challenging Parfit's interpretation of his favorite authors, or the tenability of his convergence argument, I have tried to point out a problem that only arises when his initial argument has gone through. For all I know, Parfit might indeed have shown that the different traditions (or at least some plausible versions of these traditions) agree on what matters, i.e., they might be deontically equivalent. But this, or so I have argued, is not enough to vindicate moral realism. The theories still differ in the explanations they give us for why those acts are right or wrong. I have described this as a case of underdetermination, taking inspiration from the philosophy of science. If that is an accurate description, Parfit's convergence argument might backfire and lay the ground for a similar anti-realist argument for the moral realm as has been conceived for the scientific realm.

Importantly, my remarks do not amount to a general argument against realist positions in ethics. Parfit's convergence argument, on which the underdetermination claim crucially depends, is far from being universally accepted and I have done nothing to defend it further. Even if the convergence argument is granted, and my suspicions about the remaining explanatory disagreements prove to be correct, several potentially successful options remain to block the anti-realist argument. I have tried to argue the case that at least three of the most promising options are not open to Parfit. However, whether the argument could be broadened to pertain also to realist views that do not share Parfit's conciliatory spirit or his moral epistemology remains to be seen.

In addition, we have only discussed underdetermination among three specific theories. Parfit thinks that those are the best versions of the most important traditions and, if he is right, the result would thus be of great importance. Yet we have to expect that other authors will have their own favorites. To provide a more general argument, we would need to show underdetermination to be a more pervasive phenomenon. Philosophers of science often advance more ambitious claims to the effect that *any* given theory is underdetermined. Going

67 Some might think that Parfit's reasons fundamentalism provides a way out: if reasons are the fundamental normative notion, we might well be satisfied with knowing *which* our reasons are. But this merely pushes back the question; we can still ask *why* these are our reasons.

forward, it would be of great interest to learn the prospects for such a pervasive kind of underdetermination in the realm of ethics.⁶⁸

I want to conclude with a more general observation. I think that the issues discussed have their root in the way Parfit conceives of the problem of disagreement. I submit that when doing metaethics, many (if not most) philosophers prefer to discuss their pre-theoretical notions and views on morality, and neglect (or put aside) the results of normative ethical theorizing. If we look at, e.g., the metaethical realism debate, it is noticeable that this debate does not build on the findings, preliminary as they may be, of normative ethical theorizing. Instead, an argument like that from disagreement is mostly posed in terms of disagreements between laymen, or between different cultures. This is in stark contrast to the philosophy of science. In the philosophy of science, it is scientific theories that get most of the attention, and it is to such theories that arguments like the underdetermination argument or the pessimistic metainduction refer. Parfit's way of thinking of the challenge in terms of disagreements among theories reveals a closer connection between these two domains. This allows us to adapt some questions to the moral realm that are being discussed in the scientific literature. And it might ultimately suggest that the whole project of reconciling rival theories in order to vindicate moral realism is not the most promising idea after all.⁶⁹

University of Bern
 marius.baumann@philo.unibe.ch

- 68 A promising starting point for such investigations could be the project referred to as the *consequentializing* of moral theories. Proponents of that project try to establish the claim that there is a deontically equivalent consequentialist alternative to any (minimally plausible) non-consequentialist theory. See, for example, Dreier, "Structures of Normative Theories," and more recently Portmore, *Commonsense Consequentialism*. If this were to be successful, and there did remain explanatory differences between those theories, it would suggest a more pervasive form of underdetermination.
- 69 I would like to thank: Claus Beisbart for invaluable help and encouragement; members of the Reflective Equilibrium Group (especially Georg Brun and Tanja Rechner) for several rounds of constructive criticism; Monika Betzler, Jamie Dreier, Gerhard Ernst, Brad Hooker, Christian List, Philip Stratton-Lake, Jussi Suikkanen, Folke Tersman, and Silvan Wittwer for insightful suggestions; and, finally, audiences at Edinburgh, Erlangen, and Utrecht for stimulating discussions, as well as the Swiss National Science Foundation for generous funding.

REFERENCES

- Alexander, Larry, and Michael Moore. "Deontological Ethics." *Stanford Encyclopedia of Philosophy* (Spring 2015). <https://plato.stanford.edu/entries/ethics-deontological>.
- Audi, Robert. "Intuition, Inference, and Rational Disagreement in Ethics." *Ethical Theory and Moral Practice* 11, no. 5 (November 2008): 475–92.
- Bergström, Lars. "Underdetermination and Realism." *Erkenntnis* 21, no. 3 (November 1984): 349–65.
- Berker, Selim. "The Unity of Grounding." *Mind* (forthcoming).
- Bonk, Thomas. *Underdetermination: An Essay on Evidence and the Limits of Natural Knowledge*. Dordrecht: Springer, 2008.
- Bortolotti, Lisa. *An Introduction to the Philosophy of Science*. Malden, MA: Polity Press, 2008.
- Boyd, Richard N. "Realism, Underdetermination, and a Causal Theory of Evidence." *Noûs* 7, no. 1 (March 1973): 1–12.
- Carrier, Martin. "Underdetermination as an Epistemological Test Tube: Expounding Hidden Values of the Scientific Community." *Synthese* 180, no. 2 (May 2011): 189–204.
- Darwall, Stephen. "Agreement Matters: Critical Notice of Derek Parfit, *On What Matters*." *Philosophical Review* 123, no. 1 (January 2014): 79–105.
- Deigh, John. *An Introduction to Ethics*. Cambridge: Cambridge University Press, 2010.
- Dietrich, Franz, and Christian List. "What Matters and How It Matters: A Choice-Theoretic Representation of Moral Theories." *The Philosophical Review* 126, no. 4 (October 2017): 421–79.
- Dreier, James. "Metaethics and the Problem of Creeping Minimalism." *Philosophical Perspectives* 18, no. 1 (December 2004): 23–44.
- . "Structures of Normative Theories." *The Monist* 76, no. 1 (January 1993): 22–40.
- Driver, Julia. *Ethics: The Fundamentals*. Chichester, UK: Wiley-Blackwell, 2007.
- Duhem, Pierre. *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press, 1954.
- Dworkin, Ronald. *Justice for Hedgehogs*. Cambridge, MA: Belknap Press of Harvard University Press, 2011.
- Harman, Gilbert. *The Nature of Morality: An Introduction to Ethics*. Oxford: Oxford University Press, 1977.
- Herman, Barbara. "A Mismatch of Methods." In Parfit, *On What Matters*, 2:83–115.

- Hooker, Brad. *Ideal Code, Real World: A Rule-Consequentialist Theory of Morality*. Oxford: Oxford University Press, 2000.
- Kitcher, Philip. "Real Realism: The Galilean Strategy." *Philosophical Review* 110, no. 2 (April 2001): 151–97.
- Ladyman, James. *Understanding Philosophy of Science*. Abingdon, UK: Routledge, 2002.
- Larmore, Charles. "Morals and Metaphysics." *European Journal of Philosophy* 21, no. 4 (December 2013): 665–75.
- Laudan, Larry. "Demystifying Underdetermination." In *Scientific Theories*, vol. 14, edited by C. Wade Savage, 267–97. Minneapolis: University of Minnesota Press, 1990.
- Laudan, Larry, and Jarrett Leplin. "Empirical Equivalence and Underdetermination." *Journal of Philosophy* 88, no. 9 (January 1991): 449–72.
- Mackie, J.L. *Ethics: Inventing Right and Wrong*. London: Penguin Books, 1977.
- Maxwell, Grover. "The Ontological Status of Theoretical Entities." In *Scientific Explanation, Space, and Time: Minnesota Studies in the Philosophy of Science*, vol. 3, edited by Herbert Feigl and Grover Maxwell, 181–92. Minneapolis: University of Minnesota Press, 1962.
- Nebel, Jacob. "A Counterexample to Parfit's Rule Consequentialism." *Journal of Ethics and Social Philosophy* 6, no. 2 (July 2012): 1–10.
- Newton-Smith, W., and Steven Lukes. "The Underdetermination of Theory by Data." *Proceedings of the Aristotelian Society* 52 (1978): 71–91.
- Otsuka, Michael. "The Kantian Argument for Consequentialism." *Ratio* 22, no. 1 (March 2009): 41–58.
- Parfit, Derek. *On What Matters*. 2 vols. Oxford: Oxford University Press, 2011.
- Park, Seungbae. "Philosophical Responses to Underdetermination in Science." *Journal for General Philosophy of Science* 40, no. 1 (July 2009): 115–24.
- Pietsch, Wolfgang. "Defending Underdetermination or Why the Historical Perspective Makes a Difference." In *EPSA Philosophy of Science: Amsterdam 2009*, edited by Henk W. de Regt, Stephan Hartmann, and Samir Okasha, 303–13. Dordrecht: Springer, 2012.
- Portmore, Douglas W. *Commonsense Consequentialism*. New York: Oxford University Press, 2011.
- Quine, W.V. "On Empirically Equivalent Systems of the World." *Erkenntnis* 9, no. 3 (November 1975): 313–28.
- . "On the Reasons for Indeterminacy of Translation." *Journal of Philosophy* 67, no. 6 (March 1970): 178–83.
- . "Three Indeterminacies." In *Perspectives on Quine*, edited by Robert B. Barrett and Roger F. Gibson, 1–16. Chichester, UK: Wiley-Blackwell, 1990.

- . “Two Dogmas of Empiricism.” *Philosophical Review* 60, no. 1 (January 1951): 20–43.
- Ridge, Michael. “Climb Every Mountain?” *Ratio* 22, no. 1 (March 2009): 59–77.
- Ross, Jacob. “Should Kantians Be Consequentialists?” *Ratio* 22, no. 1 (March 2009): 126–35.
- Ross, W. D. *The Right and the Good*. Oxford: Clarendon Press, 2002.
- Sayre-McCord, Geoffrey. “The Many Moral Realisms.” *Southern Journal of Philosophy* 24, no. S1 (Spring 1986): 1–22.
- . “Moral Realism.” *Stanford Encyclopedia of Philosophy* (Spring 2015). <https://plato.stanford.edu/entries/moral-realism>.
- Scanlon, T. M. *Being Realistic about Reasons*. Oxford: Oxford University Press, 2014.
- . “How I Am Not a Kantian.” In Parfit, *On What Matters*, 2:116–39.
- Setiya, Kieran. Review of *On What Matters*, by Derek Parfit. *Mind* 120, no. 480 (October 2011): 1281–88.
- Smith, William. Review of *On What Matters*, vol. 2, by Derek Parfit. *Journal of Moral Philosophy* 11, no. 2 (2014): 241–44.
- Stanford, Kyle. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford: Oxford University Press, 2006.
- . “Underdetermination of Scientific Theory.” *Stanford Encyclopedia of Philosophy* (Spring 2016). <https://plato.stanford.edu/entries/scientific-underdetermination>.
- Suikkanen, Jussi. *This Is Ethics: An Introduction*. Chichester, UK: Wiley-Blackwell, 2014.
- Tännsjö, Torbjörn. *Understanding Ethics: An Introduction to Moral Theory*. Edinburgh: Edinburgh University Press, 2002.
- Tersman, Folke. “From Scepticism to Anti-Realism.” Unpublished manuscript, 2018.
- . *Moral Disagreement*. Cambridge: Cambridge University Press, 2006.
- Tulodziecki, Dana. “Epistemic Equivalence and Epistemic Incapacitation.” *British Journal for the Philosophy of Science* 63, no. 2 (June 2012): 313–28.
- Van Fraassen, Bas C. *The Scientific Image*. Oxford: Oxford University Press, 1980.
- Wolf, Susan. “Hiking the Range.” In Parfit, *On What Matters*, 2:33–57.
- Worrall, John. “Underdetermination, Realism and Empirical Equivalence.” *Synthese* 180, no. 2 (2011): 157–72.