



In Search of the Trinity: A Dilemma for Parfit's Conciliatory Project

Marius Baumann¹

Accepted: 15 January 2021 / Published online: 4 June 2021
© The Author(s) 2021

Abstract

I outline a dilemma for Derek Parfit's project to vindicate moral realism. In *On What Matters*, Parfit argues that the best versions of three of the main moral traditions agree on a set of moral principles, which should make us more confident about the prospects of truth in ethics. I show that the result of this Convergence Argument can be interpreted in two ways. Either there remain three separate and deontically equivalent theories or there remains just one theory, the Triple Theory. Both interpretations fail to deliver what Parfit is looking for. The first interpretation leads to a situation of underdetermination of theory choice that gives rise to a skeptical challenge. The second interpretation jettisons Parfit's Conciliatory Project, that is, the reconciliation of the three moral traditions. The dilemma, I contend, is the result of Parfit failing to resolve two antithetical lines of thought. His search for the Trinity of moral theorizing must thus fail.

Keywords Derek parfit · On what matters · Convergence argument · Moral underdetermination · Moral skepticism

1 Introduction

Widespread disagreements about moral matters have for a long time caused headaches to moral realists, that is, those who think that moral discourse typically aims at making true statements about the world and at least sometimes succeeds in this quest.¹ For how could we have confidence in the claim that some moral claims are true, if we are constantly and systematically disagreeing about them? If, like Derek Parfit, one also thinks that the failure of moral realism gives us reason to think that nothing matters, the threat of disagreement

¹This follows the definition of moral realism by Sayre-McCord (2017). Realists often add that the statements are about moral facts which are in a relevant sense *mind-independent*, but since the issue of mind-independence does not play a role in what follows, I am glossing over this.

✉ Marius Baumann
marius.baumann@lrz.uni-muenchen.de

¹ Fakultät für Philosophie, Wissenschaftstheorie und Religionswissenschaft, Lehrstuhl V für Praktische Philosophie und Ethik, Ludwig-Maximilians-Universität München, Schellingstr. 10, 2. Stock, 80799, München, Germany

becomes truly grave. Facing this threat, Parfit comes up with a solution that is equally ambitious as it is ingenious. In his 2011 *On What Matters* (short: *OWM*) Parfit argues that despite what the better part of philosophers have thought for the longest time, the best versions of three of the most prominent traditions of moral theorizing - Kantianism, consequentialism, and contractualism - can be shown to arrive at the same conclusions about *what matters*; i.e., they agree about which acts are right or wrong, mandatory, permissible, or forbidden. The way he argues for this surprising conclusion is by means of what he calls the *Convergence Argument*. If successful, Parfit wagers, this argument not only puts to bed the longstanding disputes in normative ethics but should also make us more positive about the prospects of moral realism.²

Unfortunately, or so I will argue, Parfit's line of reasoning leads to a dilemma, since there are basically two ways to interpret the result of the Convergence Argument. On the first interpretation we end up with three separate theories, a Kantian, a consequentialist, and a contractualist one, which all entail the same deontic verdicts. On the second interpretation we end up with just one theory, the so-called *Triple Theory*, which somehow combines or epitomizes what's best about all three traditions. The problem is that neither interpretation delivers what Parfit is looking for. The first interpretation, as Dietrich and List (2017) and Baumann (2018) have recently suggested, leads to a problem of underdetermination of theory choice that gives rise to its own skeptical challenge. The second interpretation, in contrast, jettisons what might be called Parfit's *Conciliatory Project*, that is, the reconciliation of the three moral traditions. That this dilemma besets Parfit's project, I conclude, is no coincidence but rather the result of him failing to reconcile two antithetical lines of thought. Parfit, to take up a comparison by Larmore (2013), seems to be in search of the *Trinity* of moral theorizing and this search must fail.

Here is the plan for the paper. I start with some background on how Parfit conceives of the problem that moral disagreement poses. Following this, I sketch the Convergence Argument and outline how it can be interpreted in two ways, one that has as a result three distinct theories, the other resulting in just one theory. Next, I argue that this leads to a dilemma since neither of the two interpretations gives Parfit what he is looking for. Finally, I offer some general observations on the fault lines in Parfit's overall project, culminating in the charge that the only way for it to succeed would be if Parfit had found something akin to the Trinity of moral theorizing, which, I am afraid, he has not.

2 Disagreements, Epistemic Peers, and Conciliation

One of the most memorable impressions one gets from reading *OWM* is the sense of urgency. At the root of Parfit's thinking seems to be a deeply felt concern about the present state of normative ethics. Disagreements are the norm in the field, whether they are about the moral status of particular actions, or about the more general principles around which

²Parfit (2011b, 486) prefers the label *non-metaphysical non-naturalist cognitivism* for his own view, but there is no doubt that, according to the given definition, he counts as a moral realist. This remains true even when Parfit later changes the label to *non-realist cognitivism*, since, pace Sayre-McCord, he defines realism as entailing a robust ontological commitment. Compare Parfit (2017, 59):

We are Cognitivists but not Realists about some kind of claim if we believe that such claims can be true, but we deny that these claims are made to be true by correctly describing, or corresponding to, how things are in some part of reality.

ethicists construct their theories. This, Parfit thinks, also impacts the field of metaethics and, more specifically, the position of moral realism, i.e., the view that moral claims purport to be true and, at least in some cases, really are true. For how could we hold on to the idea of such truths if we are unable to agree on them? Since Parfit (2011b, 426-430) also holds that the only alternative to realism is a nihilism according to which nothing matters, the stakes, for him, are maximally high.

Many of Parfit's realist peers will balk at this way of describing matters. Looking at what bothers them will be instructive to get a better grip on Parfit's general outlook. First, Parfit's fear of nihilism might be uncalled for. Even many anti-realists admit that the failure to arrive at truths in ethics does not entail that nothing matters.³ Parfit's fear in this regard might thus be exaggerated.

Second, one might also think that the threat to moral realism itself is overstated. Disagreements, so one could argue, are the rule rather than the exception of our epistemic lives, and people aren't losing their epistemic faith in large numbers. Why should we think that moral disagreements speak against there being a truth of the matter if we don't accept the same conclusion about disagreements in other domains?

Part of the story why Parfit thinks that the case of moral disagreement is worse has to do with *who* is disagreeing when it comes to morality. Disagreements in ethics don't only exist among lay people but they persist between those who have most thoroughly considered matters. In Parfit's own words:

Those with whom we disagree may be responding to the same evidence, their judgment in other cases may have been as reliable as ours, and they may not be more likely to have been misled. Such people we can call our *epistemic peers*.⁴

It should be obvious how such disagreement is more threatening than disagreement between just anybody. Many disagreements among lay people can be shrugged off by citing factors such as biases, misinformation, insufficient consideration, and so on. This, or so one would hope for the discipline, is less likely to be the case when it comes to professional ethicists.

Still, one might ask, what is it about disagreements that is so threatening in the first place? Even if we can not agree about all moral matters, are we not entitled to hold on to our own beliefs at least as long as we have not been presented with better arguments to the contrary? Parfit thinks we are not. Here is how he continues the quote from above:

When we disagree with such people [*viz. peers*], we should ask whether we have sufficient reason to assume that it is our beliefs that are more likely to be true. If we conclude that we have no such reasons, we could not rationally keep these beliefs.

This is the crux of the matter. As Smith (2014, 243-244) (2011b, 426-430) points out, the foregoing quote is an expression of a *strong conciliationist view* about disagreement. Conciliationists hold that absent a good explanation of why I myself am more likely to be right than a peer, the warrant for my own belief is impacted. Parfit's version of conciliationism is a particularly strong one because he not only thinks that this should lead one to lower one's credence but he thinks that one would be irrational to hold on to one's belief.⁵ In Parfit's

³Compare Schroeder (2017) and Street (2017) for this point.

⁴Parfit (2011b, 428).

⁵Feldman and Warfield (2010) contains several articles that deal with these issues from a general epistemological viewpoint.

view, continuing peer disagreement which can not be plausibly explained away, should lead one to give up one's beliefs.

I propose that it is against the backdrop of this general epistemological view that Parfit's project in OWM should be understood. Conciliationism is not just one feature of Parfit's view. Instead, as Smith suggests, it might be the driving force behind the way Parfit goes about pursuing his overall project:

[...] supposing that the consequentialist Parfit of *Reasons and Persons* believed that he had Kantian and contractualist peers, reconciliation might be better than globalized claims of misunderstanding that he makes against his meta-ethical opponents. Given that others have found the reconciliationist methodology wanting, we may wonder about its motivations, and this strong conciliationism seems like one possible source.⁶

Elaborating on Smith, we can plausibly conjecture that after the publication of *Reasons and Persons*, Parfit takes on a meta-perspective. He looks at the state of normative ethics and realizes that there are still deep disagreements between theorists. These concern him, since he thinks that they threaten moral realism. However, Parfit is also convinced that the rival theorists are his peers, not more likely to have made mistakes than himself. He thus sees no point in simply arguing that they are all wrong but realizes that he must come up with a different solution. This sets the bar very high for his future endeavors. Yet Parfit does not shy away from the challenge. Instead, he embarks on one of the most ambitious philosophical projects in recent times: to prove that proponents of the main traditions of moral theories can actually agree about what matters. Call this his *Conciliatory Project*.

3 The Convergence Argument and two Interpretations

3.1 The Convergence Argument

The way Parfit goes about arguing is as follows. Over the course of several chapters and through an in-depth analysis of objections and counter-objections, he first identifies what he considers to be the best versions of Kantianism, consequentialism, and contractualism. Parfit (2011a, 339 and 369) is forthright about his goals here. He does not attempt to uncover the most exegetically accurate version of these theories. Instead, he is searching for the best possible shape they could take, or, more precisely, the main principles that those best versions would yield. Next, Parfit puts forward what he calls the *Kantian Argument for Rule Consequentialism*. That argument is too detailed to be considered in full length here, but what it boils down to is the following.⁷ The best version of Kantianism tells us to prefer those principles that everyone can rationally will. The best version of rule-consequentialism tells us to prefer those principles whose universal acceptance would lead to the best results from an impartial view, that is, the optimific principles. Yet the principles that everyone can rationally will are the same as the optimific ones. Kantianism, Parfit informs us, therefore *implies* rule-consequentialism. What's more, Parfit (2011a, 411-412) further argues that it is very likely that the only principles that everyone can rationally will are the ones that no one

⁶Smith (2014, 244). Compare also Darwall (2014, 82) for a similar assessment of how Parfit's epistemological views about disagreement motivate his overall project.

⁷For a detailed discussion of that argument see Otsuka (2009) and Suikkanen (2009).

can reasonably reject. Since the latter principles are those that Parfit's preferred version of (Scanlonian) contractualism recommends, contractualism is also shown to converge on the same principles. Taking all those steps together, Parfit calls this the *Convergence Argument*. If successful, it shows that, contrary to lore, three of the main traditions of moral theorizing, interpreted in the right way, arrive at the same principles about which acts are right or wrong.

The argument is not without presuppositions. Of especial importance are Parfit's views about reasons and rationality. In particular, Parfit (2011a, 377-379) makes some contentious claims about the weight of different kinds of reasons. In his mind, we often have both partial as well as impartial reasons. However, Parfit thinks that the partial reasons never outweigh the impartial ones. In addition, Parfit holds that everyone has reasons to want the optimistic outcomes and that other, non-optimific, reasons do not outweigh the optimistic reasons.

Several critics have taken issue with these views about rationality and reasons.⁸ However, controversial as these views may be, I am willing, at least for the sake of argument, to grant their correctness. Other critics have further doubted that Parfit's preferred versions of the three traditions are genuine members of those traditions. For example, Scanlon (2011, 121 ff.) argues that Parfit's Kant does not have much to do with the historical Kant. This is due to the fact that Parfit changes the direction of priority between rationality and reasons. For Parfit's Kant, we can rationally will something only if we have sufficient reason to do so. Rationality is understood through a (independently given) notion of a reason. However, as Scanlon sees it, this gets things backwards since Kant always proceeds from rationality to reasons. Thus, as Scanlon (2011, 121) describes it: "[...] there is one sense in which none of these views is Kantian [...]" Other critics have made similar charges, to the effect that one of Parfit's preferred theories is not representative of its respective tradition.⁹ I am going to set these concerns aside as well.

For the rest of this paper I will assume that Parfit has indeed shown that the best versions of the three rival traditions agree on a set of principles and thus on their verdicts about all particular cases. The problem I want to point out is thus not that the Convergence Argument fails. It is rather that *even if* the argument succeeds, it does not help Parfit's case. To see why, we need to consider next what the upshot of the Convergence Argument is supposed to be.

3.2 Two Interpretations

Somewhat surprisingly, that is not at all clear. In particular, it is not clear *how many* theories we are facing if Parfit's argument goes through. Whereas some commentators have thought that we end up with three theories, others believe that it is only one. This, I will argue, is not just of exegetical interest but it points to the deeper problems of Parfit's overall project. Let us thus look at the two interpretations in more detail.

3.2.1 Interpretation 1: Three Theories

At the end of Volume One of OWM, Parfit introduces a metaphor that arguably expresses the quintessence of his project. Summarizing the result of the Convergence Argument, he states:

⁸Compare Otsuka (2009) and Setiya (2011).

⁹Compare Morgan (2009, 59), Herman (2011, 83-84), and Larmore (2013, 668 ff.).

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.¹⁰

The idea of a group of mountaineers climbing a mountain from different sides is highly suggestive of what I take to be the first interpretation of the result of the Convergence Argument. As I understand the metaphor, Kantians, contractualists, and consequentialists start on different sides of the foot of a mountain. The space between their starting points signifies that they do not agree on what matters morally. Through modifications of their theories over time, they make progress, each taking different routes along to the summit of the mountain. Only at the top, when their theories are perfected, do they meet. They come to realize that their theories actually arrive at the same conclusions about what matters. More precisely, the theories agree on a set of principles about which acts are right or wrong. The result of the Convergence Argument is thus the insight that the most plausible versions of three of the most important moral traditions are equivalent in their deontic principles and verdicts.

This interpretation seems to be quite widely held among commentators. Dietrich and List express it in the following terms:

A striking suggestion of extensional equivalence can be found in Derek Parfit's (2011) book *On What Matters*. Parfit argues that his favorite versions of consequentialism, Kantianism, and Scanlonian contractualism essentially coincide in their recommendations and can be seen as attempts to climb the same mountain from different sides.¹¹

The notion of extensional equivalence is further taken up by Baumann (2018). Other commentators, though not using that terminology, also talk about a multitude of theories when they outline the result of Parfit's Convergence Argument. Suikkanen (2009, 2) writes: "the best versions of the main moral theories come to the same convictions about what matters", and Larmore (2013, 667) states: "[w]hen these theories are properly formulated, they each yield as rationally justified the same set of moral principles."

There is thus one prominent line of interpretation according to which, after Parfit's Convergence Argument has gone through, we are presented with three theories, each the supreme version of a prominent moral tradition, that all happen to agree about what should be done from a moral point of view.

3.2.2 Interpretation 2: The Triple Theory

Notwithstanding the wide acceptance of the first interpretation, there is also a second interpretation that has some adherents as well as some traction in the text. According to this second interpretation, the Convergence Argument results in only one theory, the so-called *Triple Theory*. The idea of the Triple Theory, just as the metaphor of the mountain, only comes into play towards the end of Volume One. Parfit (2011a, 412) informs us that if the Convergence Argument is sound, Kantianism, contractualism, and consequentialism can be *combined* in the following way:

¹⁰Parfit (2011a, 419). The manuscript that Parfit circulated among his peers prior to the publication of OWM even referred to the metaphor in its title; "Climbing the Mountain".

¹¹Dietrich and List (2017, 425)

Triple Theory: An act is wrong if and only if, or *just when*, such acts are disallowed by some principle that is

- (1) one of the principles whose being universal laws would make things go best
- (2) one of the only principles whose being universal laws everyone could rationally will
and
- (3) a principle that no one could reasonably reject.

The Triple Theory, Parfit goes on to explain, describes

[...] a single *higher-level* wrong-making property, under which all other such properties can be subsumed, or gathered.¹²

We will look at the Triple Theory in more detail when we consider the dilemma. For now, it is only relevant to notice that these remarks suggest a different interpretation of the result of the Convergence Argument. Instead of three, we are presented with only one theory. The three theories which arrive at the same principles are somehow combined, leading to one super-theory that picks out what is common to all lower-level wrong-making properties.

Several commentators accordingly speak of only one theory. Herman, when summarizing the result of the Convergence Argument, speaks of a *hybrid* theory:

Derek Parfit's *On What Matters* offers an avowedly hybrid theory of morality, or at least of the part of morality that tells us which acts are wrong.¹³

Chappell similarly describes the goal of Parfit's project as follows:

The omega-point of such convergence would be to find the single complete and objectively correct theory of the right, or to get as close as we can to that single theory.¹⁴

We thus find (at least) two widely held interpretations of what the results of the Convergence Argument are.

4 The Dilemma

It is difficult to assess which interpretation is more commonly held and even more difficult to assess which has a better claim to be considered Parfit's final interpretation (if he had one). Luckily, for our purposes, we don't have to make such an assessment because, as I am about to argue, both fail to vindicate Parfit's overall goal of rescuing moral realism. Instead, Parfit faces a dilemma: no matter which interpretation is correct, his overall project is bound to fail.

¹²Parfit (2011a, 414). The fact that Parfit, in the preceding quote, formulates the Triple Theory using the phrase "if and only if, or *just when*" might at first look confusing, since he otherwise uses the phrase "if and only if, and because" when defining moral theories. However, this quote makes clear that the Triple Theory is also about *wrong-making* properties and therefore about *why* some acts are wrong, not just *which* acts are wrong. I take up this point again in Section 4.2.

¹³Herman (2011, 83).

¹⁴Chappell (2012, 170-171).

4.1 Horn 1: Underdetermination and Skepticism

Let us begin with the problem that arises for the first interpretation. It has its origin in a simple observation. Even though Kantians, consequentialists and contractualists may agree on *which* acts should be done, they still disagree about *why* those acts should be done. As one commentator puts it:

[...] even if it turned out to be the case that consequentialists can genuinely accommodate commonly held moral beliefs about, for instance, our having special duties to our family members and other people close to us, a claim of which I am suspicious, they would nevertheless find themselves obliged to say so for the wrong reasons.¹⁵

Morgan frames the problem in terms of an explanatory shortcoming of consequentialist theories. However, the point can be put in a more general (and less partisan) way. Dietrich and List (2017) do so by drawing a distinction between, on the one hand, the body of action-guiding verdicts a theory yields, and, on the other hand, the theoretical explanation of why these are the correct verdicts. Slightly more technically, they distinguish between a *rightness (or permissibility) function*, which yields a set of action-guiding verdicts for each situation and a *reasons structure*, which encodes the underlying reasons that are being given for why specific acts are right or wrong, permissible, mandatory or forbidden.¹⁶ The permissibility function thus answers the question ‘Which actions are right or wrong?’, whereas the reasons structure answers the question ‘Why are those actions right or wrong?’.

Moral theories, Dietrich and List insist, include *both* a rightness function and a reasons structure. Yet this, crucially, opens up the possibility that theories might agree when it comes to the former, but not the latter. They may agree on which acts are right or wrong but still differ when it comes to their accounts of why this is so. Such theories are, in other words, *deontically equivalent* while at the same time being *theoretically incompatible*.¹⁷ This, Dietrich and List go on to explain, is structurally analogous to a phenomenon from the philosophy of science, the so-called *underdetermination of theory by evidence*. As philosophers of science have discussed following Duhem (1906) and Quine (1953), in the the history of science it has often been the case that more than one theory was able to account for all the available data.¹⁸ What is more, these theories often made incompatible claims about unobservables. They were, as it is sometimes put, *empirically equivalent* while at the same time being theoretically incompatible. Taking both the deontic content of moral theories as well as the empirical content of scientific theories to be the extension of those theories, we can see that there is a structural analogy between science and ethics: in both realms, theories can be extensionally equivalent while being theoretically incompatible.¹⁹

The analogy is highly interesting in its own right, but for the present purpose, we are interested only in how it affects Parfit’s project. Dietrich and List don’t help us much further here, since they don’t consider the metaethical repercussions of their observation about

¹⁵Morgan (2009, 20). Compare also Suikkanen (2014, 104).

¹⁶To facilitate reading, I will use the short-form *right or wrong* to refer to all the deontic verdicts in what follows.

¹⁷Baumann (2018) provides a lengthy defense of the claim that Parfit’s preferred theories are indeed theoretically incompatible. I take up this issue in Section 4.2.

¹⁸Famous examples being the Copernican and the Ptolemaic view of planetary motions in the sixteenth century or particle and wave theories of light in the nineteenth century.

¹⁹Dietrich and List further claim that the same structural analogy can be found in the so-called *consequentializing* debate. See also Baumann (2019) for an in-depth investigation of that claim.

moral underdetermination. However, Baumann (2018) takes up that task and argues that the analogy is problematic for Parfit. The reason for this is that, starting with Duhem (1906), philosophers of science have argued that underdetermination poses grave problems for scientific realism. Yet that means that if Parfit is correct, and we are facing a structurally analogous situation in ethics, there arise similar problems for realist positions in ethics. To show how, Baumann construes the following argument, which mirrors the structure of similar arguments in the philosophy of science:

- 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.²⁰

Before we look at the premises in more detail, a first preliminary point to note is that the argument is by no means meant as a general argument in favor of a skeptical position in metaethics. Several of the premises are highly controversial, as Baumann is quick to acknowledge. Instead, the argument is custom-tailored to counter Parfit's use of his own Convergence Argument to vindicate moral realism. A second preliminary point is that the argument does not presuppose that because there is a similarity in the *nature* of the empirical and the deontic, what is true for the empirical must also be true for the deontic. Instead, the argument simply makes use of a structural analogy. If a very specific constellation regarding scientific theories and their extension gives rise to skeptical worries in science, then the exact same situation in ethics should give rise to skeptical worries, as well.

What about the premises then? Baumann's discussion is too detailed for us to consider at length here, but a few comments are in order. The critical notion in P1 is that of the evidence of a moral theory. A narrow construal of the evidence would only include the deontic consequences of a moral theory, i.e., the particular deontic verdicts it yields. However, on such a construal, P1 is not very plausible, since non-deontic factors, such as simplicity, non-adhocness, etc., can arguably make a theory more believable than a deontically equivalent counterpart. For P1 to be more plausible, we would thus have to assume that the rival theories also fare equally well when it comes to non-deontic factors. Parfit does not tell us much about such non-deontic factors of his preferred versions of the moral traditions, so it is unclear what his position is on this matter. However, we don't need to know the answer, since the dilemma persists either way. If the three theories are indeed equally believable,

²⁰Baumann (2018, 206). Compare Kukla (1998, 58), Psillos (1999, 164), and Douven (2008, 294-295) for similar arguments in the philosophy of science.

considering all possible evidence, then P1 is correct. If they are not, and one is more believable than the others, then we should arguably prefer it over the others and we are left with only one supreme theory when the Convergence Argument has gone through. In that case, I refer the reader to the discussion of the second horn. P3 I consider to be relatively uncontroversial, at least for realists. P2 and P4 are both more controversial, but they are direct correlates of Parfit's strong conciliationist position in epistemology, which entails that if we don't have sufficient reason to assume that our beliefs are more likely to be true than those of our rivals, we can not rationally hold on to them. Finally, the last premise, *UMT*, is short for *underdetermination of moral theories*. If we grant that the main traditions of moral theorizing are putting forward incompatible explanatory claims, *UMT* is just a restatement of what follows from the Convergence Argument according to the first interpretation.

Since the conclusion of the argument is that we should not consider our best moral theories to be true, Parfit has not bettered the stance of moral realism. On the contrary, the opposite seems true. Mere disagreement, even if it is between our best theories, is something most moral realists can live with, assuming that we might one day be in a better position to decide between the theories. Underdetermination, however, forestalls that option. It entails that there is no way for us to resolve the remaining explanatory disagreements. For a strong conciliationist, that means we have to give up belief in the theories.

But maybe this is too fast. As Parfit is well aware,

Intuitionists need not claim that, in ideal conditions, these disagreements would all be completely resolved. (Parfit (2011b, 548))

Indeed, demanding a full consensus on all moral questions seems uncharitable. If Parfit's Convergence Argument is successful, a considerable amount of disagreement has been resolved. The only disagreement that remains is about explanation. Shouldn't this be enough? Couldn't Parfit take the position that even if the explanatory disagreements remain, the comprehensive extensional agreement that has been reached is enough to vindicate intuitionism and, with that, his form of realism?

I am highly skeptical that this reply can be made to work, for at least four reasons. First, disagreements about the fundamental explanatory claims of our theories are not just any kind of disagreement. Even though we cannot demand from Parfit that he resolve all moral questions, he himself acknowledges that:

[Intuitionists] must defend the claim that, in ideal conditions, there would not be deep and widespread moral disagreements. (Parfit (2011b, 548))

Note that this claim is much stronger than what many other moral realists claim. Realists often claim that, in order to believe that we have the ability to know moral truths, it is enough to agree on some (or most) moral issues, and the debate between moral realists and moral antirealists is often framed in the way that the former claim that there are *some* moral truths whereas the latter deny that there are *any* such truths. However, these realists are also not bothered by persisting peer disagreements in the way that Parfit is. Parfit explicitly claims that we need to be able to settle our deepest moral disagreements in order to be justified to believe in moral truths. Yet the explanatory disagreements between the main traditions are some of the deepest, if not *the* deepest, disagreements in moral theorizing.

Second, if we think about it, we soon realize that it is not just a disagreement about one claim which is at issue. We have so far restricted our observations to the fundamental explanatory claims alone, because this is what Parfit focuses on. However, no moral theorists would simply put forward one explanatory claim and think that the topic of explanation is thereby settled. Instead, moral theorists typically provide us with a full story about why

it is the outcomes, or universalizability, or an idealized agreement, which render a specific course of action morally right. Parfit has not reached agreement on any part of these stories. The remaining disagreements are thus about the whole *class* of explanatory statements.

Third, and relatedly, there is something very ad-hoc about the idea that we shouldn't worry too much about the explanatory disagreements. If extensional disagreements are dangerous enough to threaten our beliefs that anything matters, why aren't explanatory disagreements? Parfit would have to explain this difference. It is one thing to claim that we cannot reach agreement about all substantive matters in ethics. It is a completely different thing if we cannot agree about a whole part of moral theorizing, explanation. The latter surely calls for an explanation, and Parfit has not provided one.

Finally, and besides the ad-hocness, there is something peculiarly disturbing about the idea that we would have knowledge about the complete extension of moral theories, but don't have any confidence in the explanation. If we don't know the explanation, then why should we trust our judgements about extension? This is decidedly different from the case where a part of the extension remains unknown. Imagine that Parfit had achieved convergence on all moral questions, except the ones regarding future generations. This would be an inconvenience, but we might still think that the remaining disagreements are due to our limited epistemic situation and thus remain sanguine about them. Parfit's case seems very different from this and not in a good way. Indeed, as I will explain in Section 5, comparison to the philosophy of science shows that restricting one's attention to the extension of moral theories, while neglecting the explanatory disagreements, is probably not what a realist should do.

In sum, I don't think that Parfit has a satisfactory answer to the skeptical problem of underdetermination that follows from the first interpretation.

4.2 Horn 2: Giving up on Conciliation

What is the problem with the second interpretation? Surely, underdetermination can not be it, since on the second interpretation we are left with only one moral theory and theory choice can only be underdetermined if there is, well, a choice. The second interpretation might also, at least on first look, fit well with Parfit's overall Conciliatory Project. For how much more conciliation can we hope for than to be shown that the seemingly incompatible theories can be combined into one?

This is exactly where the problem starts, however. Until now, I have not said much about how this combination into one theory is supposed to work in detail. Yet the additional information that Parfit provides lays the ground for the second challenge. Here is how Parfit describes the way that consequentialism is integrated into the Triple Theory:

Though this view [the Triple Theory] is Consequentialist in its claims about which *principles* we ought to follow, it is not Consequentialist either in its claims about *why* we ought to follow these principles, or in its claims about which *acts* are wrong. This view, we might say, is only *one-third* Consequentialist.²¹

As Parfit explains, the Triple Theory is not wholly consequentialist, but only one-third. The other parts are, presumably, Kantian and contractualist. We might therefore think that everyone has been given their due share, making conciliation possible.

²¹Parfit (2011a, 418).

However, as Parfit presents it in the quote above, a consequentialist should not accept the offer. The reason for this, once more, has to do with the explanatory dimension of moral theories. Moral theories are not mere algorithms to produce the correct verdicts. They also explain to us why these are the correct verdicts. In Dietrich and List's terminology, moral theories do not only consist of a rightness function but they also have a distinct reasons structure. If, as Parfit explains in the quote above, the Triple Theory is not consequentialist in its claims about why we ought to follow the specified principles, then the consequentialist reasons structure has fallen out completely from the Triple Theory.

The question then becomes whether we can still say that a theory is consequentialist, even just partly, if it does not include a consequentialist explanation? I think the answer, clearly, has to be negative. Consequentialism is a broad tent, encompassing a wide variety of theories, as anyone familiar with the debates within this tradition is (painfully) aware of. Act consequentialists urge us to look directly at the outcomes of specific acts, with some focusing on the actual and others focusing on the expected consequences. Rule consequentialists instead urge us to consider which rules will yield the best outcomes, be it if the rules are universally satisfied or if they are universally accepted, as Parfit's preferred version has it.²² In addition, consequentialists differ in their theories of the good, that is, what they ultimately take the value of outcomes to consist in.²³ All these differences do not decide whether one is a consequentialist or not. For what ultimately unites consequentialists, is that in their explanations of what morality demands from us, they will make reference to outcomes. The constitutive feature of all consequentialist theories, thus, is that they refer to outcomes in their explanations of what we ought to do (or which rules we ought to follow).

As I see it, the fact that consequentialists evaluate actions or rules according to their contribution to the best outcomes is what *makes* them consequentialists. It is the answer to the question "Why is an act (or rule) right or wrong?" that lets one see the true colors of a moral theorist. Since the Triple Theory does not accept consequentialists' answer to this question, it is not *one-third* consequentialist, but rather not at all consequentialist. The Triple Theory might have the correct rightness function, but it has the wrong *reasons structure* to count as a consequentialist theory.

If readers have doubts about this claim, they may ask themselves what their intuitions are in a case where a consequentialist happens to arrive at just some of the same verdicts as a Kantian. Would we in that case assume that said consequentialist is to some degree a Kantian (and the Kantian is to the same degree a consequentialist)? I think it is clear that this does not follow. The two theorists simply agree on what should be done in some cases. Their explanations for why we should act this way can still be very different. Yet if this is so for particular cases, then why should it be different for whole theories? The Triple Theory is not consequentialist to such an extent as it arrives at the verdicts that some consequentialists may also arrive at. It would only be consequentialist if it put forward a consequentialist explanation of why these are the correct verdicts.

The same goes, *mutatis mutandis*, for any tradition of moral theorizing. Parfit only tells us that the Triple Theory is not consequentialist in its claim about why we ought to follow

²²Parfit does consider act consequentialism more closely in OMW 3 and argues that some versions of act consequentialism might also be closer to the other traditions than previously thought. However, I agree with Hooker (2020) that Parfit ultimately rejects act consequentialism. Compare Parfit (2017, 413-416 and 433-435) for his reasons to reject act consequentialism.

²³Finally, even when they agree on a theory of the good, consequentialists might still disagree because they might be maximizers, or prioritarrians, or sufficientiarrians. They might thus agree on *what* the good is, but not on how it should be weighed.

the principles outlined. Thus we don't know whether it is Kantian or contractualist (or something entirely new) when it comes to explanation. Whatever it is, however, the implication is that the other one of these traditions also does not find its explanatory claim incorporated in the Triple Theory. Since I don't think that either Kantians or contractualists constitute the majority of moral theorists, the Triple Theory does not incorporate the explanatory claims of a majority of moral theorists. This, to be sure it not a problem for the Triple Theory *per se*. The Triple Theory might be a regular nonconsequentialist theory, and if it also turns out to be the best theory, then so much worse for consequentialism (and all the nonconsequentialist alternatives). The problem is rather that, so understood, the Triple Theory does not deliver what Parfit is looking for. For now we have in effect given up on reconciliation. The Triple Theory does not tell everyone that they are right, instead it tells a majority of theorists (all those who don't find their fundamental explanatory beliefs expressed in the theory) that they are wrong. What is more, they are not just wrong about some parts of their theories that might easily be fixed. Instead they are mistaken about their fundamental claims about what makes acts right or wrong. That sure does not resemble any claim of reconciliation.

But maybe I am misunderstanding the nature of the Triple Theory. Maybe the Triple Theory does not specify one fundamental explanatory claim, but instead specifies three *partial* explanations. In other words, the theory claims that if an act violates an optimific principle that is *one* way it is wrong, if it doesn't follow from a principle that everyone could rationally accept this is *another* way it is wrong, and if the action doesn't follow from a rule that nobody could reasonably reject, that's the third way it is wrong. The full reason for the wrongness of an act is thus the combination of all three claims, all claims are necessary parts of the full explanation but none of them is a sufficient explanation. Thought of in this way, the Triple Theory might be considered *one third* consequentialist (and Kantian and contractualist) after all and would thus deserve to be called an attempt at reconciliation. Does this understanding help Parfit to avoid the second horn of the dilemma?

I don't think so. Neither do I think that it is what Parfit had in mind, nor that the proposal can be made to work. My reason for doubting that this is how Parfit understood the Triple Theory is twofold. First, as we saw in the quote at the beginning of this section, Parfit explicitly tells us that the Triple Theory is *not* consequentialist in its claims about why we should follow the specified principles. This passage would thus have to be considered a mistake and considering its prominent place in the book, I can't believe Parfit would have made such a mistake. Second, the idea that moral theories only specify partial grounds of rightness and wrongness certainly isn't the commonly held view, neither is the view that they specify *multiple* partial grounds. If the Triple Theory was supposed to only present us with partial claims of wrongness, wouldn't Parfit have told us this explicitly, especially since it is decidedly not how most ethicists usually understand moral theories?

Ultimately, I cannot exclude the possibility that this was Parfit's actual understanding of the Triple Theory, but I find it very unlikely. Be that as it may, however, I don't think that this understanding is of any help, since it has grave problems of its own. For what would it mean that the theory gives us three partial explanations? Several options come to mind, but none seem promising. First, the explanations are not on different levels. The case is not akin to, e.g., an explanation of the workings of our minds according to a psychological model versus a physiological model, or like a micro-economic versus a macro-economic explanation. Clearly, all three explanations are on the same level, that of normative ethical theorizing, and are thus *rival* accounts. Second, it is not the case that the three explanations all pick out different sets of principles and only the principles on which they agree are the correct ones. Instead, each of them picks out the same set of principles. Third, the explanations do not specify principles for different realms of the moral, which only together combine to a full

moral view. It isn't like one of them only specifies principles for one realm, for example the realm of what we owe each other, and the others supplement this with principles for other realms. The explanations thus don't combine to one unified explanation, but instead they overdetermine the principles. This is not logically impossible, but it is definitely far away from how we normally understand moral explanation. If presented with a consequentialist explanation of an act's wrongness, Kantians typically don't reply 'That's right, but don't forget that the action could also not be willed to become a universal law'. Rather, Kantians insist that their explanation is the only correct one even if the consequentialist happens to agree with them on what to do.

My impression is that the Triple Theory tacitly plays on an ambiguity between a merely extensional and an explanatory understanding of moral theories. If moral theories only specified which acts are right, it would not be a problem if they overlapped. We could think of them as merely different ways of systematizing a body of principles. The problem only arises when we think of the theories as also offering explanations. The reason for this is that explanations, due to their intensional character, can differ even if they are (necessarily) co-extensive.²⁴ At least this is what those of us think, who don't see moral theories as mere mechanisms to produce deontic verdicts. Parfit could of course insist that the Triple Theory is a merely extensional theory and he could go on to claim that the only thing that matters about moral theories is their extension. That, however, would be a truly revisionary position and I see not indication that he was ready to go there.²⁵ It is also less than clear that it would help Parfit's case with his peers, to say the least.²⁶ I therefore do not think that this understanding of the Triple Theory provides a feasible option to steer clear of the second horn.

Let me be try to be as clear as possible as to where the force of the objection lies. Certainly, even considering Parfit's strong conciliationism, we can not expect him to effectively convince everyone, or else explain his project to have failed. Disagreement is only threatening when our peers *justifiably* hold on to their own views. They can not simply insist on their own views, if presented with good reasons to change them. Thus, if Parfit offers convincing arguments, we must expect his peers to accept them, even if that means that they have to considerably modify their theories. One's opponent's strong conciliationism is no invitation to simply stick to one's guns.

However, if what I have said so far is correct, Parfit's solution does not only ask his peers to partially modify their moral views, e.g., adapt their theory of the good. Instead, it asks the majority of them to give up their foundational explanatory claim. If, as I have further argued, these foundational claims are constitutive of the different moral traditions, Parfit is effectively asking his peers to abandon their preferred moral traditions. That is, he is not only asking them to make changes *within* their preferred theoretical framework but to effectively leave that framework. That surely is not in the spirit of his Conciliatory Project. To claim that the majority of theorists has been wrong about their most fundamental

²⁴ Compare Berker (2018) on this point.

²⁵ Remember the quote from Section 3.2.2 where Parfit explains that the Triple Theory describes a *wrong-making* property.

²⁶ Parfit would have to accept what Jamie Dreier calls:

Extensionality Thesis (short: *ExT*): ExT says that nothing but extension matters in a moral view. (Dreier (2011, 98))

Yet even Dreier (2011, 114) himself acknowledges that his Extensionality Thesis is *more controversial* than the thesis about the deontic equivalence of theories.

explanatory convictions bears very little resemblance to an attempt at conciliation. Even if we are very optimistic about the persuasive power of Parfit's argument and the readiness of his peers to change their views, it simply does not seem reasonable to assume that he can convince them to give up their most dearly held explanatory beliefs. The Triple Theory, contrary to its conciliatory name, asks the best part of moral theorists to give up their story of what makes acts right or wrong. That, I think, does not deserve to be considered an attempt at conciliation. Thus while the second interpretation avoids the threat of underdetermination it jettisons Parfit's Conciliatory Project.

5 The Search for the Trinity

5.1 Two Irreconcilable Lines of Thought

If what has been said so far is correct, then no matter which interpretation of the result of the Convergence Argument is correct, it does not fit together with Parfit's overall project. This, I want to argue in the remainder of the paper, is no coincidence. The fact that there is a confusion about the number of theories that the Convergence Argument leads to is not due to a mere lack of clarity in Parfit's presentation. It is rather a direct consequence of Parfit failing to reconcile a deep tension in his thinking.

One way to get at this tension is to consider the discrepancy between, on the one side, the set up and much of the rhetoric of Volume 1 and, on the other side, the proposed solution at the end, involving the Triple Theory. Parfit's mood, throughout OWM, is an irenic one. Fear of skepticism gets him started and removing the disagreements which give rise to skepticism is what motivates his project. Starting from the point of strong conciliationism, one would expect Parfit to try and convince as many people as possible. That is indeed the prevalent impression one gets through much of the book. As it unfolds, one feels the pull on all theorists to come together and modify their theories towards deontic convergence. The metaphor of the mountain, I have argued, epitomizes that mood. At the end, when we reach the summit of the mountain, we find out that everyone was attempting to reach the same end point, just following a different path.

However, at around the same stage in the book, the idea of the Triple Theory makes its entrance. Parfit informs us that the end product of his argument is this new hybrid theory. This move, considering the rest of the book, comes as quite a surprise. As Parfit himself puts it earlier in the book

[...] rather than proposing a new moral theory, I shall try to learn from some existing theories, hoping to get somewhat closer to the truth. (Parfit (2011a, 174))

One starts to wonder: if Parfit simply wanted to present us with his own superior theory, then why take the detour? Why get everybody on board if it turns out that they were wrong after all? For the message of the Triple Theory is not 'You have all been right' but 'Most of you have been wrong'. This is at odds with the conciliationist appeal which otherwise dominates the book. Thus the switch from the Three-Theory interpretation to the One-Theory interpretation also makes for a confusing change in mood.

In addition, there is also a confusing switch in perspectives taking place here. Parfit, I have argued, starts from a kind of meta-perspective. Why is it, he wonders, that philosophers just as knowledgeable as himself and not more likely to have committed mistakes, arrive at very different conclusions? When he adapts this perspective, it seems that Parfit considers it to be futile to simply put forward another first-level theory. What is needed is a new

kind of solution that takes into consideration the fact that ethicists don't seem to be able to convince their peers by simply putting forward their own theories. Yet the solution involving the Triple Theory is on the level of first-order theory-building. Notwithstanding its lofty name, Parfit, by putting forward the Triple Theory, joins the ranks of those who have, in his own estimation, unsuccessfully tried to convince everyone else that they are the only ones who are right. The Triple Theory is just another theory, proposed by another philosopher, with low chances of being accepted by his peers, especially those who don't find their explanatory convictions upheld in the end result.

Note that I emphatically do not claim that realists should cease to improve their preferred moral theories in order to make it more likely that they turn out to be true. There is nothing wrong with reacting to peer disagreements by proposing one's own theory (or someone else's) to be superior and thus more likely to be true. There is nothing wrong *per se* with globalized claims of misreadings, for that matter, we often find them in philosophy. The problem for Parfit is that his starting point is from a meta-level that already precludes such a strategy. He looks at the field of normative ethics, sees persistent peer disagreement, and explicitly states that he has no reason to assume that his peers are more likely to be wrong than he himself. I take it that much of the enthusiasm that Parfit's project undoubtedly generates stems from him framing it as if he were not going to make the standard move of presenting his own, allegedly better, theory, but were instead trying to reconcile the conflicting parties. Looked at from this perspective, the Triple Theory has to be seen as a disappointment. Had Parfit started the book openly stating that he was going to construct his own superior theory, I doubt very much it would have produced the same excitement.

There are, I think, two lines of thought that Parfit's project fails to reconcile. On the one hand, there is the thought, possibly inspired by Parfit's intimate knowledge of the dialectical situation in normative ethics, that we won't make progress if ethicists continue to wield charges of global misunderstandings and mistakes against each other. On the other hand, there is his fear of nihilism which makes it necessary to solve the remaining disagreements. Yet these thoughts call for different approaches. The first thought calls for conciliation and acceptance of other ethicists' views. The second thought necessitates a decision for one of the theories. Parfit can not have it both ways.

Comparison to the philosophy of science yields some insights into what might have gone wrong here. In the philosophy of science, it is *anti-realist* leaning philosophers who have for a long time tried to convince us that our best theories have extensionally equivalent rivals. The purpose of establishing extensional equivalence, very basically, is to prevent realists from deciding between the rival theories on the basis of the data, and thus make them less confident in their beliefs. Realists, in comparison, can welcome extensional disagreements between theoretically incompatible theories since that leaves open the possibility that they might one day be able to decide between the theories on the basis of extension. The idea of a crucial experiment, for example, is exactly to bring to the fore such differences.

Thought of in this light, Parfit's idea of helping realism by proving the best theories to be deontically equivalent becomes increasingly puzzling. By resolving the extensional disagreements, Parfit is in effect robbing his realist peers of the means to decide between the theories. No trolley case, no other crucial (thought)-experiment will help us if theories agree on their extensions. Parfit's strategy thus has an oddly anti-realist flavor.

Perhaps vaguely aware of this, Parfit might then have introduced the idea of the Triple Theory as a possible solution. If there remains only one theory, there is no more need to resolve disagreements on the basis of extensional differences. That is also what a scientific realist would try to do. However, the scientific realist would arguably not see the need to frame this as a conciliatory approach. The realist's aim, very roughly, is to falsify the

incorrect alternatives, not to reconcile them with the correct theory. Realists also do not need to bother about other people having been wrong, as long as we come to the correct conclusions. Considering this, it might start to appear like Parfit's longing for conciliation, and the strong conciliationism that leads him to it, themselves have a skeptical bent. I think that there is something to be said for this. Scientific realists typically tend to take a more steadfast view of disagreements.²⁷ The fact that there are competing scientific theories does not bother them much, since they assume that there is a fact of the matter independently of what they know about it. It is anti-realists who are much more impressed by disagreements and who urge us to lessen our beliefs if we can not reconcile our disagreements.²⁸

5.2 The Triple Theory and the Trinity

I think that the idea of the Triple Theory has rightfully come to be seen as the epitome of Parfit's project. However, what it epitomizes are the unresolved contradictions in Parfit's project. It is rather telling, then, that Larmore compares it to another notoriously puzzling idea:

Parfit's Triple Theory is much like the Trinity: in the three-in-one, one of the three enjoys a priority over the other two.²⁹

Larmore, to my knowledge, doesn't introduce the comparison to the Trinity as a way of illustrating a shortcoming in Parfit's reasoning. Yet the comparison evokes the foregoing dilemma in a very vivid way. On the one hand, we have the three-in-one - God the father, the son, and the holy spirit. On the other hand, we have God the father reigning over the others. But which is it, the skeptics and unbelievers will ask, one or three entities?

This same structural mystery haunts Parfit's Triple Theory. In order for the Triple Theory to achieve the conciliation Parfit is looking for, it needs to incorporate all three of the main traditions. That is why it (allegedly) consists of one third of each tradition, the three-in-one. However, since normative moral theories are also about explanation, and the explanatory claims of the three traditions are incompatible, one of those traditions' explanation needs to be prioritized. Just as in the Trinity God the Father reigns supreme, only one tradition of moral theories can have it right when it comes to the fundamental question of what makes an act right or wrong. If we don't want to give up the explanatory ambitions of moral theorizing, we have to make up our minds. However, in that case it is no longer clear how the Triple Theory could count as a three-in-one theory at all. The analogy to the Trinity is thus highly suggestive of the dilemma at the heart of Parfit's project.

Not surprisingly, and just like the Trinity, the idea of the Triple Theory leads to lots of confusions. This is nowhere more obvious than in the fact that there isn't even agreement about which moral theory reigns supreme. Larmore (2013, 668) thinks that the Kantian element of the theory is more basic and accordingly speaks of *Kant the Father*. However, others seem to disagree. Herman (2011, 83) holds that: "[...] the hybrid theory winds up having a strongly consequentialist cast", whereas Scanlon (2011, 138) thinks that contractualism wins the day. Each of the three commentators seems to assume that one of the traditions

²⁷Compare Feldman and Warfield (2010, 4) for a similar assessment.

²⁸This might go some way in explaining the impression of some commentators that Parfit's fear of disagreement somehow fits uneasily with other aspects of his philosophy, especially as concerns metaethics. Compare Larmore (2013, 667) and Darwall (2014, 83) for this point.

²⁹Larmore (2013, 668).

must have a privileged position in the final result, but the fact that it remains unclear which tradition that is, does not speak to the clarity of the idea of a Triple Theory.

Attempts at making more sense of the talk of the Triple Theory don't always help, either. As a first example, consider this short paragraph:

At the heart of Parfit's book is an argument for convergence in moral theory, according to which the most plausible version of Kantian ethics is a form of contractualism that is not only consistent with, but in fact entails, the most plausible version of consequentialism.³⁰

Setiya's depiction is very much in sync with how Parfit himself talks about his project and it might seem innocent at first look. However, what exactly is it supposed to mean that the Kantian theory *entails* consequentialism? I do have an understanding of what it means that the Kantian theory entails the same principles as a consequentialist theory does. This is probably what Suikkanen (2009, 18) has in mind when he summarizes the result of OWM as follows: “[w]e have then argued from Kantian premises to a rule-consequentialist conclusion.” However, the principles are not all that constitutes the consequentialist theory. In order to entail the whole theory, the Kantian theory would also have to entail consequentialist *explanatory* principles. This can not be the case, however, or else we would have incompatible explanatory principles combined within one theory.

As a second example, consider how Parfit (2011a, 413) talks about the *triplely supported* principles. The idea here, presumably, is that the principles are triplely supported because they are entailed by each of the three traditions. However, as we have seen, the Triple Theory does not accept the consequentialist explanation of *why* these are the correct principles. The consequentialist explanation is wrong. Yet, one might ask, how can an explanation that is wrong *support* justification of the principles? Also, if those principles are triplely supported, that is arguably by three distinct theories. That, however, fits better with the first interpretation than the one involving the Triple Theory. The Triple Theory, if it is to be one theory, can not support its own principles threefold.

One might finally think that the two interpretations aren't incompatible after all, because the Triple Theory is not a theory on the same level as the original Kantian, contractualist, and consequentialist ones. Instead, it could be considered some kind of meta-theory. This reading gets some backing from Parfit's claim, quoted above, that the Triple Theory identifies a *higher-level* wrong-making property. If the Triple Theory picks out a higher-level property, it might itself be considered a higher-level theory, leaving intact, at a lower level, the three theories that arrive at the same conclusions about what matters and thus provide independent support for the principle. That, however, only doubles our problems. For now we are left with the three disagreeing theories on the lower level and the internally inconsistent theory at the higher level.

The Triple Theory, I conclude, is much like the Trinity; an idea that obscures more than it illuminates. What it obscures is that it is supposed to do double-duty. On the one hand, it has to do justice to the legacy of all three major rival traditions. On the other hand, it has to solve the problem of disagreements. However, since the rival traditions are just that, *rivals* when it comes to their explanatory claims, the Triple Theory can not do both and thus has to fail.

³⁰Setiya (2011, 1281).

Acknowledgements I am grateful to Claus Beisbart, Monika Betzler, Georg Brun, Jamie Dreier, Gerhard Ernst, Brad Hooker, Christian List, Tanja Rechitzer, Philip Stratton-Lake and several anonymous referees for their valuable feedback, as well as audiences at Edinburgh and Utrecht Universities for helpful discussions.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baumann M (2018) Parfit, convergence, and underdetermination. *Journal of Ethics and Social Philosophy* 13(3):191–221
- Baumann M (2019) Consequentializing and underdetermination. *Australas J Philos* 97(3):511–527
- Berker S (2018) The unity of grounding. *Mind* 127(507):349–365
- Chappell T (2012) Climbing which mountain? a critical study of derek parfit, on what matters. *Philos Investig* 35(2):167–181
- Darwall S (2014) Agreement matters: critical notice of derek parfit, on what matters. *Philos Rev* 123(1):79–105
- Dietrich F, List C (2017) What matters and how it matters: a choice-theoretic representation of moral theories. *Philos Rev* 126(4):421–479
- Dreier J (2011) In defense of consequentializing. In: Timmons M (ed) *Oxford studies in normative ethics* (1), Oxford University Press, pp 97–119
- Douven I (2008) Underdetermination. In: Psillos S, Curd M (eds) *The routledge companion to philosophy of science*, Routledge, pp 292–301
- Duhem P (1906) *La Theorie Physique: Son Objet et sa Structure*. Marcel Riviera et Cie. English Translation by Philip P. Wiener, *The Aim and Structure of Physical Theory* Princeton University Press [1954]
- Feldman R, Warfield TA (2010) *Disagreement*. Oxford University Press, Oxford
- Herman B (2011) A mismatch of methods. In: *On what matters, volume two*, Oxford University Press, pp 83–115
- Hooker B (2020) Parfit's final arguments in normative ethics. In: McMahan J, Campbell T (eds) *Essays in Honour of Derek Parfit: Normative Ethics and Personal Identity*, Oxford University Press, pp 1–21
- Kukla A (1998) *Studies in Scientific Realism*. Oxford University Press, Oxford
- Larmore C (2013) Morals and metaphysics. *Eur J Philos* 21(4):665–675
- Morgan S (2009) Can there be a kantian consequentialism?. In: Suikkanen J, Cottingham J (eds) *Essays on derek parfit's on what matters*, Wiley-Blackwell, pp 39–60
- Otsuka M (2009) The kantian argument for consequentialism. In: Suikkanen J, Cottingham J (eds) *Essays on Derek Parfit's On What Matters*, Wiley-Blackwell, pp 61–78
- Parfit D (2011a) *On What Matters. Volume One*. Oxford University Press, Oxford
- Parfit D (2011b) *On What Matters. Volume Two*. Oxford University Press, Oxford
- Parfit D (2017) *On What Matters. Volume Three*. Oxford University Press, Oxford
- Psillos S (1999) *Scientific Realism: How Science Tracks Truth*. Routledge
- Quine WvO (1953) Two dogmas of empiricism. *Philos Rev* 60(1):20–43. Reprinted in *From a Logical Point of View*
- Sayre-McCord G (2017) Moral realism. In: Zalta EN (ed) *The Standford Encyclopedia of Philosophy*. Fall 2017 edition
- Scanlon TM (2011) How I Am Not a Kantian. In: *On what matters, Volume Two*, Oxford University Press, pp 116–139
- Schroeder M (2017) *What Matters About Metaethics?*. In: Singer P (ed) *Does anything really matter?*, Oxford University Press
- Setiya K (2011) Review of Derek Parfit, 'On What Matters'. *Mind* 120(480):1281–1288

- Smith M (2009) Two kinds of consequentialism. *Philos Issues* 19(1):257–272
- Smith W (2014) Book review: On what matters, volume 2. *Journal of Moral Philosophy* 11(2):241–244
- Street S (2017) Nothing ‘Really’ matters, but that’s not what matters. Singer, P. editor, *Does Anything Really Matter?* Oxford University Press
- Suikkanen J (2009) Introduction. In: Suikkanen J, Cottingham J (eds) *Essays on Derek Parfit’s On What Matters*, Wiley-Blackwell, pp 1–20
- Suikkanen J (2014) *This is Ethics: An Introduction*. Wiley-Blackwell

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.