CHAPTER 3

Parfit and the Case Study of Case Studies

On what I have dubbed the *Textbook View*, theories of alternative traditions disagree about both what makes acts right or wrong as well as which acts are right or wrong. Despite this arguably representing the mainstream view in normative ethics, there has also been a vocal minority that demurs, arguing that the traditions might not be as different as is commonly assumed. The most ambitious expression of this to date is due to Derek Parfit. In the first book of his opus summum, *On What Matters*, Parfit argues, over the course of several hundred pages, that the most plausible versions of three of the most important traditions of moral theorizing actually agree on *what matters*. They converge on the same set of principles about which acts are right or wrong. Parfit himself is very optimistic about what this result entails, evoking the metaphor of climbers meeting at the summit. Yet, or so I will argue, the different routes these climbers take are of the utmost importance. What we end up with on top of the mountain are three incompatible theories arriving at the same set of verdicts, which is properly described as a case of moral underdetermination.

I start the chapter with some historical background. Parfit's argument is the latest in a succession of attempts to bring the moral traditions closer together, which arguably spans at least from J. S. Mill up to one of Parfit's own teachers, Richard Hare. Still, Parfit's attempt is much more ambitious, which is why I focus on it for the rest of the chapter. I outline how Parfit arrives at his surprising conclusion via his *Convergence Argument*. I then proceed to interpret the results of this argument by first arguing against two interpretations that are suggested in Parfit's writing, the *Conciliatory* and the *Triple Theory* interpretations. Both I find lacking. Instead, I set out why we should interpret Parfit's project as a case study of moral underdetermination.[1]

[1] Some of the material of this chapter appears in Baumann (2018) and Baumann (2021a).

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

3.1 Doubting the Dogma

Any mainstream view in philosophy reliably attracts heretics, and so it has been with the Textbook View and its antagonistic picture of the moral traditions. To provide some perspective on Parfit's project, I thus want to start with a quick look at two philosophers who preceded Parfit in challenging the Textbook View.

The first figure is J. S. Mill. The question of how the rival moral traditions relate to each other appears in the context of his discussion of the just and the useful in his 1871 book *Utilitarianism*. Mill (1871, pp. 87 ff.) reflects on the fact that considerations of justice have often served as counter examples to utilitarianism. A well-known example is of a judge who has the option to unjustly sentence an innocent person to death in order to pacify a mob and thereby avert the killing of several innocents. Utilitarianism, it is held, cannot account for our conviction in such cases that sentencing innocent people is against what justice demands.

This reasoning seems wrong to Mill. Although he acknowledges that the two ideas – justice and utility – are distinct, he holds that they actually point in the same direction when we consider a longer period of time:

> [...] [T]he Just must have an existence in Nature as something absolute – generically distinct from every variety of the Expedient, and, in idea, opposed to it, though (as is commonly acknowledged) never, in the long run, disjoined from it in fact. (Mill, 1871, p. 87)

Considerations of justice and utility will thus converge, Mill thinks, and, remarkably, he even thinks that this fact is commonly acknowledged.

Mill (1871, p. 97) then goes on to comment on Kant more specifically and, again, finds more convergence than one might suspect. He argues that Kant would have to agree with him that the idea of justice, correctly understood, presupposes that everyone profits from it. To see why, Mill turns to Kant's discussion of the Categorical Imperative (CI). Mill argues that Kant could not have made sense of the CI if not with a utilitarian interpretation in the back of his mind. He illustrates this for the universal law formulation. In Mill's opinion, we have to give the universal law formulation a utilitarian interpretation, unless we want to render it meaningless:

> To give any meaning to Kant's principle, the sense put upon it must be, that we ought to shape our conduct by a rule which all rational beings might adopt *with benefit to their collective interests*. (Mill, 1871, p. 97)

The reason for this is that barring an appeal to overall utility, nothing in the CI would stop rational people from accepting egoistical maxims.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

Since Kant clearly wants to exclude egoistical maxims, his only option, on Mill's understanding, is to opt for an interpretation of the CI that is based on considerations of utility. Mill, it appears, is trying to get Kant on board with the utilitarian enterprise by attributing to him certain utilitarian arguments. Mill considers this to be a charitable interpretation of Kant, since the latter's views would come out as incoherent or meaningless unless they are further supported by considerations of utility.

Did Mill's reasoning convince others that we should think of Kant as much closer to utilitarianism than we (and Kant himself) might at first sight think? Not if we ask Richard Hare. Taking up Mill's discussion of Kant's relation to utilitarianism,[2] Hare (1997, p. 148) maintains that it has become something of a dogma to position Kant and the utilitarians on opposite sides of the moral spectrum. Yet, in an eponymous article, Hare asks: "Could Kant have been a utilitarian?" He answers in the affirmative, his argumentative strategy being basically twofold. Wherever he can, Hare interprets Kant's remarks in a utilitarian vein. Wherever this does not work, Hare rejects Kant's remarks as being an unfortunate consequence of his rigoristic upbringing.[3]

This strategy is based on a distinction that Hare (1997, p. 148) draws between, on the one hand, Kant's formal theory, drawn primarily from different formulations of the Categorical Imperative, and, on the other hand, Kant's substantive claims and judgments, as evidenced by his examples. Making use of this distinction, Hare tries to prove that Kant's theory is actually compatible with utilitarianism. Hare (1997, pp. 152–153) argues, for example, that the *mere means* formula is in no conflict with utilitarianism as long as we exclude duties toward ourselves as manifestations of Kant's rigorism. The same goes for the formula of universal law. Hare (1997, pp. 153–155) thinks that the formula can be made compatible with utilitarian thinking as long as we specify the maxims in sufficiently sophisticated ways. The fact that Kant includes simple maxims like "Thou shalt not break promises" is, again, dismissed as a consequence of his upbringing. Hare (1997, p. 154) sees no inconsistency in people rationally accepting that some cases of promise-breaking would be universalized. Kant's argument, according to which we cannot will a law that allows for the breaking of promises since this would render the practice of promise giving untenable, strikes him as weak. Echoing Mill, Hare instead argues that Kant's remarks

[2] Hare (1997, p. 148) acknowledges Mill's contribution.

[3] Compare Hare (1997, pp. 148 and 154–155).

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

only become understandable when explicated by means of utilitarian thought.

Hare (1997, p. 148) softens the paper's controversial title by admitting that Kant himself was no utilitarian and would arguably not have been convinced by utilitarian reasoning, due to his acquired moral sensibilities.[4] Contrary to what the title suggests, Hare's interest is thus not in whether the historical Kant might have been a utilitarian but rather in whether his theory can be made compatible with utilitarian sensibilities. Moreover, making a distinction between Kant's theory and his concrete judgments is certainly not unheard of. Few philosophers today who consider themselves Kantian feel obliged to accept all of Kant's judgments in order to stay true to the framework. Still, the argument that Kant's system of morality is actually compatible with utilitarian conclusions is certainly provocative.[5]

Mill and Hare are but two examples of philosophers challenging the antagonistic picture encoded in the Textbook View. There have been many others.[6] Yet none has been as thorough and extended as Derek Parfit's *On What Matters*.[7] In a veritable tour de force, spanning several hundred pages, Parfit argues that the best versions of three of the most important families of moral theories, namely Kantianism, consequentialism, and contractualism, arrive at the same conclusions about what matters. While Mill and Hare don't do much more than gesture in the direction of agreement, Parfit attempts to show how we might arrive there via a very detailed and intricate succession of arguments. Parfit also departs from the spirit of many of his predecessors. In order to achieve convergence, both Mill and Hare disregard many of Kant's views in a wholesale way. Their arguments have a clear consequentialist bent, and they are unlikely to find much uptake among Kantians. Parfit is much less dismissive of non-consequentialist views. His aim is not to make Kantians (and contractualists) see the

[4] Hare (1997, p. 147) also clarifies that the goal of his article is to ask a question, not to answer it.

[5] A somewhat positive assessment can be found in Cummiskey (1990) and Forschler (2013); critics include Timmermann (2005) and Kalokairinou (2011).

[6] The theorist who is most obviously missing in this line is Sidgwick. I do not treat his *Methods of Ethics* in detail for two reasons. First, what he says about the relation between different theories is more extensive than both Mill's and Hare's comments and hence cannot be treated with such brevity. Second, the discussion of Sidgwick is complicated by the fact that he includes egoism as a third major tradition which, as Crisp (2020, p. 270) explains, is considered even less of a serious contender today than it was in Sidgwick's day. In contrast, Parfit's choice of theories is much more relevant to today's discussion. Another, more recent, potential case study is Cummiskey (1996). I will say more about his *Kantian Consequentialism* in Chapter 5 but have to omit it here for sake of space.

[7] Indeed, both Hooker (2010) and Singer (2011) attest to the more general importance of *On What Matters* when they call it the greatest/most significant work of ethics since Sidgwick's *Methods of Ethics*.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

light of consequentialism. Instead, he aims to show that Kantians (and contractualists), while staying true to their traditions, can modify their theories so as to arrive at the same conclusions about what matters. Finally, the motivation behind Parfit's project is also very different from Mill's and Hare's. Their interest in the convergence of moral theories seems to be motivated primarily by two desires: First, to clear up what they take to be mistaken views that lead to an unnecessarily antagonistic picture of the moral traditions; second, to strengthen their own consequentialist theories by showing that other traditions could, and indeed should, go their way too.

3.2 Climbing the Mountain

The way Parfit goes about searching for convergence between the main traditions is as original as it is ingenious. First, over the course of several chapters and through a rigorous analysis of problems and objections, Parfit identifies what he considers the best versions of Kantianism, consequentialism, and contractualism. Parfit (2011a, p. 339 and p. 369) frankly acknowledges that his main interest here is not in staying true to every detail of the traditions' original shapes. Instead, he is searching for the most plausible forms they could take. Standing on the shoulder of giants, he attempts to make more progress.

In Kant's case, this means searching for Kant's supreme principle. Parfit (2011a, pp. 177–342) considers a multitude of candidates in Kant's writing, tirelessly addressing objections and modifying Kant's original ideas. In the process, he rejects many prominent Kantian ideas, for example, the *mere means* principle or the *dignity* or *respect* principles, for either leading to wrong verdicts or failing to guide us at all.[8] Ultimately, Parfit (2011a, pp. 338–342) settles on the universal law formulation, albeit a modified version thereof. The problem with the idea of universalization, Parfit holds, is that Kant thinks of it from the perspective of single agents. However, like Mill and Hare before him, Parfit sees no inconsistency in agents preferring a principle that would only benefit themselves or people relevantly similar to them. For example, men might conceivably have no problem with a patriarchal society. Parfit's solution is to build impartiality into the principle.[9] Instead of a principle that focuses on what agents can

[8] For an illuminating discussion, see Suikkanen (2009b, pp. 9 ff.).

[9] Compare Morgan (2009, pp. 44–45) for this reading of Parfit. Parfit (2011a, pp. 289–300) also gives up completely on Kant's notion of a maxim.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

rationally will from the first-person perspective, Parfit (2011a, p. 342) favors the following principle:

> "Everyone ought to follow the principles whose universal acceptance everyone could rationally will."

A similar attentiveness and love of detail leads Parfit to a modified version of his favorite contractualist principle. Parfit (2011a, pp. 351–355) first rejects Rawlsian contractualism because it comes too close to (act-)utilitarianism. Instead, Parfit (2011a, pp. 360–370) prefers Scanlon's version, though, again, with reservations. Whereas Scanlon initially proposes his own principle as only pertaining to the class of actions that concern what we owe to each other, Parfit wants a more encompassing principle that applies to all morally relevant acts. This brings Parfit (2011a, p. 369) to the following principle:

> "An act is wrong just when such acts are disallowed by some principle that no one could reasonably reject."

Finally, Parfit (2011a, pp. 370–403) considers consequentialism. After going through a series of arguments and objections, he opts for rule- instead of act-consequentialism. This is highly significant in its own right, as we shall see. Furthermore, it is also interesting in terms of Parfit's own philosophical development. Parfit (1984) himself had earlier been much closer to an act-consequentialist theory. Since act-consequentialism is often considered to be a more revolutionary theory than rule-consequentialism, Darwall (2014, pp. 80–81) accordingly identifies a shift in Parfit's thinking from a more *antiestablishment*-leaning earlier phase to a more *conservative*-leaning later phase. Perhaps aware of this, in Volume 3 of *On What Matters*, Parfit returns to act-consequentialism and suggests that it, too, might be closer to the other three traditions, as well as less at odds with common-sense morality, than most people think. Still, Parfit ultimately rejects act-consequentialism (for reasons too detailed to be repeated here), and in what follows I will accordingly focus on his original arguments involving rule-consequentialism.[10] The version of rule-consequentialism that Parfit ultimately opts for is itself quite classical. In Parfit's understanding, rule-consequentialists first identify the *optimific* principles, that is, the principles which would have the impartially best outcome if they were to be accepted by everyone. They then stipulate that everyone is supposed to follow those

[10] Compare Parfit (2017a, pp. 413–416 and pp. 433–435) for his reasons to reject act-consequentialism. See also Hooker (2020) for an excellent discussion of these reasons and Skorupski (2018) and Stangl (2020) for Parfit's other arguments relating to act-consequentialism.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

principles. Parfit (2011a, p. 377) accordingly suggests that the best version of consequentialism entails the following principle:

> "Everyone ought to follow [the] optimific principles."

Many of Parfit's moves here are as original as they are controversial. The heart of the argument is yet to follow, though. In a remarkable twist, Parfit (2011a, p. 379) next construes what he calls the *Kantian Argument for Rule Consequentialism.*[11] The main idea, roughly, is this: Those principles that everyone can rationally will are simply those that, if universally accepted, would make things go best in the impartial sense, that is, the optimific principles. 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 rule-consequentialism. The argument relies very heavily on substantial views about reasons and rationality, which Parfit had spent a whole separate Part of *On What Matters* defending. In particular, Parfit (2011a, pp. 377–379) makes a contentious claim about the weight of different kinds of reasons. In his mind, we often have both partial as well as impartial reasons. However, he thinks that the impartial reasons are always at least *sufficiently* weighty so as not to be outweighed by the partial ones. In addition, regarding these impartial reasons, 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 and that these reasons are at least not decisively outweighed by non-optimific considerations.[12] This is what ultimately allows him to argue that Kantians would indeed choose the same principles as consequentialists. Having thus argued that Kantianism and consequentialism are compatible, Parfit (2011a, pp. 411–412) further argues that the only principles that everyone can rationally will are also highly likely to be the ones that no one can reasonably reject. This tops things off, since it grants compatibility with Parfit's preferred version of (Scanlonian) contractualism as well.

Taking all those steps together, we can dub this the *Convergence Argument.*[13] If successful, it shows that, interpreted in the right way, three of the main traditions of moral theorizing arrive at the same principles. These

[11] For a detailed discussion of that argument, see Otsuka (2009) and Suikkanen (2009b).

[12] The argument is actually more complex. Parfit thinks that there are objective truths about the relative strengths of partial and impartial reasons. However, these truths are, even in principle, very imprecise. We should thus assume that very often we have sufficient reasons for both.

[13] Parfit also uses this name for the more restricted argument that shows that Kantianism and Scanlonian contractualism can agree. However, since this use is misleading considering that the two also converge with rule-consequentialism, I will use the locution to refer to Parfit's overall argument for the convergence of all three traditions.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

principles are *deontic principles* in the sense that they specify the deontic status of classes of acts. Since, very plausibly, verdicts in particular cases follow directly from these principles, the theories must also agree on the former. Thus, although Parfit does not use this terminology, we can note that the Convergence Argument leads to theories that are *extensionally equivalent*. Parfit does not tell us in detail what the extension or the deontic content of these principles is, other than that the principles are the optimific ones. However, in Volume 3 of *On What Matters*, Parfit (2017a, p. 434) informs us that the principles he had in mind are those of common-sense morality. This is a significant addendum since it does not follow directly from the outlined arguments. It entails that Kantianism, contractualism, and consequentialism actually agree with common-sense morality on a set of principles about what we should do.

Parfit's Convergence Argument has already attracted a great deal of scrutiny and criticism. Some critics take issue with the possibility of convergence itself, coming up with moral choice situations where it seems that Parfit's preferred theories do not lead to the same deontic verdicts.[14] Others have doubted that the versions of the theories Parfit works with are genuine members of the respective moral traditions.[15] Still others have made charges of partiality, to the effect that Parfit, contrary to his expressed conciliatory approach, prefers one of the traditions to the others.[16] I will come back to some of these, but I cannot consider them here. In this, I find myself in the same shoes as philosophers of science who cannot scrutinize all scientific claims and thus have to base their views at least partly on preliminary results from the sciences. What I do want to take issue with is what Parfit takes the results of his own Convergence Argument to imply for our understanding of the relation between the moral traditions.

3.3 Failed Conciliation and Moral Underdetermination

Parfit's own remarks in *On What Matters* suggest at least two ways in which we could interpret the results of his Convergence Argument. Both, I will argue, turn out to be unsuccessful, prompting me to propose an alternative interpretation in terms of underdetermination.

[14] Compare Ross (2009, pp. 145 ff.), Herman (2011, pp. 84 ff.), and Chappell (2012, pp. 174 ff.).

[15] Compare Scanlon (2011, pp. 121 ff.), Morgan (2009, p. 59), Herman (2011, pp. 83–84), and Larmore (2013, pp. 668 ff.).

[16] Compare Herman (2011, p. 83), Larmore (2013, p. 668), and Scanlon (2011, p. 138).

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

The Conciliatory Interpretation

Following the conclusion of the Convergence Argument, Parfit ends Volume One of *On What Matters* on a memorable note:

> 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. (Parfit, 2011a, p. 419)

The metaphor of the mountain epitomizes what might be called the *conciliatory interpretation* of Parfit's project.[17] According to this interpretation, Parfit is neither offering a new moral theory, nor is he taking sides. Instead, the three traditions are on the way to the same conclusions, albeit they have started from different assumptions. We can appreciate the conciliatory spirit of this view if we compare it to Parfit's predecessors. Mill and Hare clearly think that it is Kant's mistakes alone which have to be rectified in order to achieve convergence. Parfit disagrees. Instead, he thinks that all three traditions are on the right track, even though they all might need some improvements here and there. The metaphor of a group of hikers meeting at the summit beautifully depicts this conciliatory perspective.

At the same time, the metaphor strikes me as particularly telling of what I see as the fundamental problem with the conciliatory interpretation. I take the metaphor to signify that when the different theorists reach the summit, that is, when they have perfected their theories, they will notice that they agree on all their verdicts. Thus, there are indeed no disagreements remaining on this level. Yet 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 top of the mountain?

To put it less metaphorically, the first interpretation fails because all that has been shown is that different theories can indeed lead to the same consequences about what we should do. Parfit might think that he has therefore settled all the relevant conflicts between those theories. However, this is not so because even 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

[17] The manuscript that was widely circulated before the publication of *On What Matters* was titled *Climbing the Mountain*. Compare also Skorupski (2018, p. 610) for the view that Parfit's project is one of *conciliation* instead of *revision*.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

point was stressed early on by Suikkanen (2014), who asks why Kantians and consequentialists, despite seemingly agreeing about particular cases, nevertheless disagree. His answer, I think, is spot on:

> 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. (Suikkanen, 2014, p. 104)

What Suikkanen is bringing to our attention here, of course, is the second function of moral theories that we encountered in Chapter 2. Moral theories are not just 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. Yet when it comes to these explanations, Parfit's preferred theories continue to disagree. Kantians claim that what makes acts right or wrong is that they (fail to) conform to principles whose universal acceptance everyone could rationally will. Contractualists claim that an act is right or wrong because it is (dis)allowed by some principle that no one could reasonably reject. Finally, consequentialists hold that an act is right or wrong because it does (not) follow from optimific principles. The theories thus pick out different right-(and wrong-) makers, and the conciliatory interpretation fails because it only partly reconciles the main traditions, leaving untouched what arguably amounts to the most fundamental difference between the traditions.[18]

This has been an admittedly quick overview, and I will have a lot more to say about the explanatory disagreements shortly and then repeatedly over the course of the book. But before doing so, l want to bring into focus the second interpretation.

The Triple Theory

Readers familiar with *On What Matters* will have missed a prominent feature in my presentation of Parfit's view so far: the *Triple Theory*. At the end of Volume One, Parfit argues that we can arrive at a Triple Theory,

[18] For similar points, compare also Bykvist (2013, p. 349), Hooker (2020, p. 7), and Chappell (2021, p. 33).

which combines what is best about all the different traditions. The Triple Theory, he explains, describes

> […] a single higher-level wrong-making property, under which all other such properties can be subsumed, or gathered. (Parfit, 2011a, p. 414)

In this passage, it does not seem as though the end result of Parfit's argument will be three different theories that agree on their verdicts. Instead, the three moral traditions can be synthesized into one theory that incorporates parts of all of them. The Triple Theory, Parfit is anxious to make clear, does not reduce the three traditions to just one of them:

> 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. (Parfit, 2011a, p. 418)

Instead, the Triple Theory is supposed to be a genuinely hybrid theory.

Does this fact make the interpretation fare better than the conciliatory one? I don't think that it does. On the contrary, the Triple Theory raises a host of problems. Some of them are exegetical, concerning how the Triple Theory fits into the rest of Parfit's project. These need not concern us here.[19] What interest us here are the philosophical problems. Most importantly, we need to ask what it means to say that the Triple Theory *combines* aspects of all three traditions. Parfit (2011a, p. 26) clearly seems to think that it is a combination of all three theories, speaking of rule-consequentialism as a *component* of the Triple Theory. But there are two ways of understanding that claim, neither of which is satisfying. When Parfit claims that the Triple Theory is not consequentialist in its claims about *why* we ought to follow the principles, he might mean that consequentialists' foundational explanatory principle is not included in the Triple Theory. If this is so, then I simply don't think that a consequentialist could or should accept the Triple Theory. If what has been said in Chapter 2 is correct, moral theories have an explanatory side as well. More strongly, it is a constitutive feature of a consequentialist theory that it explains the deontic status of acts by reference to their outcomes. The Triple Theory, on this reading, would not include such an explanatory claim. As Chappell (2021, p. 33) suggests, its explanatory contribution would thus be only incidental, the actual wrong-making being done by another theory. Can we still say that the theory is at least partly consequentialist under such circumstances? I don't think so.

[19] I go into the details of this in Baumann (2021a).

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

On this reading, the Triple Theory is not *one-third* consequentialist; it isn't consequentialist at all.

Alternatively, one might assume that the Triple Theory, contrary to the quote above, does after all include the consequentialist explanatory principle *along* with the foundational principles of the other theories. This possibility is mentioned by Hooker:

> These three theories disagree with one another about what the one unifying foundational principle is. But these three theories' disagreement about what the unifying foundational principle is doesn't keep the three theories from being the elements of Parfit's Triple Theory. (Hooker, 2020, p. 7)

On this reading, the Triple Theory includes three foundational explanatory principles that pick out different grounds of right-making. Yet, as Hooker himself states, that would mean that the Triple Theory includes incompatible parts, and I am not sure that there is a more obvious reason to reject a theory.

In the end, the whole talk of the Triple Theory being only in part consequentialist, contractualist, and Kantian is deeply perplexing. It is probably no coincidence, therefore, that the analogy Larmore comes up with is to the Trinity:

> 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. (Larmore, 2013, p. 668)

Needless to say, this is not a flattering comparison for the Triple Theory. Including parts of three *incompatible* theories is no advantage for a theory, but rather a decisive reason to reject it.

Summing up, both the conciliatory and the Triple Theory interpretations have deep problems.[20] If an alternative interpretation does not share these problems, it would certainly have a significant advantage.

The Underdetermination Interpretation

Enter the underdetermination interpretation. According to it, Parfit's preferred theories arrive at the same conclusion about what we should do, yet they still offer different accounts of why we should do so. This avoids both problems of the aforementioned interpretations. It does not claim that the different traditions have been fully reconciled, as the conciliatory

[20] In addition, as I have argued in Baumann (2021a), the combination of the conciliatory and the Triple Theory interpretations leads to a dilemma for Parfit's overall project that is unlikely to be resolved.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

interpretation does; nor does it presuppose that all the traditions have been combined in some sort of hybrid theory, as the Triple Theory does. Instead, just as scientific theories can sometimes be extensionally equivalent while at the same time remaining explanatorily incompatible, so too can moral theories.

That Parfit's project can be interpreted by way of such an analogy to the philosophy of science was first noted by Dietrich and List (2017). They observe that:

> 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. (Dietrich and List, 2017, p. 425)

They further contend that their *reason-based representation* of theories:

> [...] supports Parfit's claim that different moral theories can in principle climb the same mountain from different sides, reaching the same action-guiding recommendations at the summit, albeit via different routes. Moreover, we can accept this general structural point, irrespective of whether we are persuaded by Parfit's own example of it: the purported convergence of consequentialism, Kantianism, and Scanlonian contractualism. (Dietrich and List, 2017, p. 451)

While remaining agnostic on whether Parfit's particular project of establishing deontic convergence succeeds, Dietrich and List thus provide indirect support for the general idea that such convergence is possible. Still, as I have noted in Chapter 2, Dietrich and List draw 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. This opens up the possibility of moral underdetermination. Theories may agree on which acts are right or wrong but still differ when it comes to their accounts of why this is so. This, Dietrich and List (2017, p. 425) observe, is structurally analogous to underdetermination in science.

Since Dietrich and List are primarily interested in the formal representation of moral theories, they do not go into the details of how to understand the explanatory claims that constitute a theory's reason structure. The grounding model can be put to use to fill this gap. As we have seen in Chapter 2, moral explanation, on this model, is about identifying the grounds of what makes acts right or wrong; it is about picking out the right- or wrong-makers. Theories from different traditions identify different grounds. In Parfit's case, one theory claims that acts are

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

wrong because they cannot be willed by everyone to become universal laws; another theory claims that they are wrong because they could be reasonably rejected; and the third claims that they are wrong because they do not lead to the optimific results. If we understand these claims in terms of picking out ultimate wrong-makers, the traditions still disagree.

Moreover, the grounding model can explain why these theories radically disagree even if they *necessarily* arrive at the same verdicts about what we should do. The reason for this is that the grounding relation is *hyperintensional* and thus distinguishes between different explanations even if they turn out to be necessarily co-extensive. Even if the principles that can be willed by everyone to become universal laws turn out to be the same as those that cannot be reasonably rejected as well as those that lead to the optimific result, the different theories consider only one of these facts to ground the rightness of the principles. On the grounding model, Parfit's preferred theories come out as radically disagreeing because they make different claims as to what is fundamentally prior in ethics. This is important since, as we learned from our discussion of scientific underdetermination in Chapter 1, whether the disagreement between two theories is radical is *the* crucial question we need to answer if we want to know whether what's at stake is indeed an interesting form of underdetermination. As our discussion of scientific underdetermination further showed, this is not a simple question to answer since scientific theories might be reconcilable or reducible to each other in a way that may not be immediately obvious. This is true for moral theories as well. Hence, it is important that we can give an account of how the theories in a case of moral underdetermination might nevertheless be incompatible when it comes to their explanations.

That Parfit's theories come out as radically disagreeing on the grounding model should not come as a big surprise. The grounding model provides a specific account for how to understand moral explanation and disagreements about such explanation. But the idea that the relation between the moral theories is an antagonistic one is inherent in the Textbook View itself. The Textbook View tells us that the moral traditions disagree about explanations *in addition* to yielding different deontic verdicts. Indeed, the moral traditions are often *defined* in opposition to each other.[21] The dialectical situation, I take it, is thus that unless we are presented with an argument to the contrary, we are justified in assuming that Parfit's versions of the rival traditions are explanatorily incompatible even if they may agree

[21] For example, Alexander and Moore (2016, pp. 2 ff.) use consequentialism as a foil to define deontology.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

on what we should do. The mere fact that Parfit might have shown that (the best) versions of these theories can agree about the deontic verdicts does not imply that the explanatory disagreements vanish as well. Instead, if we wanted to claim that the explanatory disagreements can be solved, that is something that would need to be shown *in addition*.

Summarizing what has been said so far, there is a very strong indication that the default understanding of the moral traditions is one of radical disagreement with regard to explanation. The grounding model provides independent support for this assumption by introducing a specific understanding of moral explanation that can make sense of the remaining disagreements in Parfit's case. Parfit, as far as I can tell, does not give us any explicit arguments why we should think that the explanatory disagreements have been resolved as well. We are thus justified, I think, in assuming that they remain unresolved.

Still, before wrapping up the chapter, let me mention one complication, which should also serve to illustrate why presupposing the grounding model isn't entirely trivial. That complication is presented by Parfit's own metaethical views. Although Parfit firmly believes in moral truths, Parfit eschews the term "realist" throughout Volumes One and Two, instead calling his own position *non-metaphysical non-naturalism*.[22] In Volume Three he makes this point even more salient by changing the label to *non-realist cognitivism*.[23] He explains the position as follows:

> 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. (Parfit, 2017a, p. 59)

According to this view, moral truths do not entail any metaphysical claims.[24] Unfortunately, Parfit does not tell us much about the alternatives. He seems content simply challenging his opponents to come up with an explanation of their *ontologically weighty* notions, while defending his own notion mostly negatively by looking for companions in guilt in mathematics and logics. But, presumably, the non-metaphysical view also pertains to moral explanatory claims. Just as we do not take on any metaphysical commitments when we attribute to acts the property of rightness, claiming that some feature makes an act right or wrong does

[22] Compare Parfit (2011b, pp. 486–487).

[23] Although the label changes, Skorupski (2018, p. 603) argues convincingly that Parfit's ontology remains broadly the same throughout the three volumes.

[24] Compare Dworkin (2011) and Scanlon (2014) for similar views.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press

not entail any metaphysical consequences. Yet, in that case, we cannot understand moral explanation in terms of grounding since the latter is a metaphysical notion.[25]

It is not the place here to inquire into the details of Parfit's metaethics. Instead, let us, for the sake of argument, assume that Parfit is right. Does this mean that the disagreements between the traditions are thereby resolved? Surely not. To see why, consider the following scenario. Imagine a situation where someone hurts someone else, and you ask me why I think this act is wrong. My answer is that the person could not have rationally wanted everyone to accept a principle that allows this act. Now, let us imagine that on another occasion, we are facing another act of hurting that is similar to the first one in all relevant respects. But this time, when asked, I tell you that the act was wrong because it did not lead to the optimific results. Would you not be confused, and rightly so? "It is either/or, you have to make up your mind!" is what you would likely reply.

What the example shows is that explanatory statements can be incompatible even if we don't construe them in a metaphysical way. Parfit, I think, would accept this. Despite having no ontological implications, Parfit (2011b, p. 479) informs us, moral claims can nevertheless be true in the strongest sense. Yet if Parfit wants truth in the strongest sense, he also has to accept disagreement in the strongest sense. Parfit clearly accepts this in another context when he claims that:

> [...] *different* theories might all be true. My claim was about *conflicting* theories. Two theories conflict when they make or imply claims which are contradictory, so that these theories cannot both be true. (Parfit, 2017b, p. 194)

As Parfit here implicitly acknowledges, cutting the metaphysical slack of the grounding model does not change anything about the fact that theories might still radically disagree. The grounding model accounts for these differences in metaphysical terms, and in this, I have argued, it is a semantically accurate depiction of much of moral theorizing. But that does not mean that the disagreements simply vanish on another model. Given what has been said above, Parfit is not justified in claiming that all differences between the alternative traditions have been resolved. Instead, he has provided us with an impressive, in-depth case study of the moral version of underdetermination.

[25] Laskowski (2018, p. 501) wonders whether Parfit, in Volume Three of *On What Matters*, had actually become sympathetic to the hyperintensional turn that led to tools like grounding. However, even Laskowski does not really see Parfit put the tool to work.

https://doi.org/10.1017/9781009492454.007 Published online by Cambridge University Press