CHAPTER 8

Looking Back and Ahead

I want to finish the book by doing two things. First, I offer two hypotheses on why we even ended up in a situation of underdetermination in ethics in the first place, and why now. Second, to round things off, I return to the Textbook View and the puzzle that set off the investigation into moral underdetermination.

8.1 How We Got Here

One fact that calls for an explanation is why the analogy between ethics and philosophy of science, as far as underdetermination is concerned, has only very recently been remarked upon by Dietrich and List (2017). Considering that the underdetermination idea has a fairly long history in the philosophy of science, one might wonder why the analogy has only been noted just now. The answer seems easy: The projects that lead to moral underdetermination are all of fairly recent descent (at least in philosophical terms). There was no analogy to make because the phenomenon in ethics came into life only recently. However, the question then becomes: Why did we get to this point at exactly this time? Is moral underdetermination a timeless phenomenon that has just not been noticed before? Or is the whole phenomenon more indicative of our present state of moral theorizing? Assessing this question in detail would require a much more thorough historical analysis than I can provide. Nevertheless, I want to suggest two hypotheses that might explain why the topic of converging theories, and with it that of moral underdetermination, has gathered speed in recent times. These are little more than guesses at this point.[1] However, if true, they also help alleviate two specific doubts about moral underdetermination that might have by now entered the reader's mind.

[1] I consider the matter in some more detail in Baumann and Beisbart (unpublished manuscript).

Advanced Dialectics

A first hypothesis points to the advanced dialectical stage that ethical theorizing has reached. The idea, roughly, is that during decades of theoretical disputes, the rival traditions have been brought more and more into line with each other extensionally as a result of an ongoing dynamic of adaptation in light of objections leveled by rival theorists.

This dynamic is especially striking in the case of consequentializing. Schroeder recounts how a succession of (alleged) improvements has changed consequentialists' theory of value over time:

> Just as Mill's utilitarianism improved on Bentham's by allowing for two kinds of pleasure and Moore's consequentialism improved on Mill's by allowing for other basic intrinsic goods, ART-ists [Schroeder's term for consequentializers] claim that their view improves on ordinary consequentialism by simply filling in a more sophisticated axiology. (Schroeder, 2007, p. 279)

Schroeder depicts consequentializers as joining the ranks of a long succession of philosophers that have tried to improve on utilitarianism. Replacing the monist theory of value we find in the earliest utilitarian theories, we see a more and more sophisticated conception of value developing from Mill to Moore. These developments allowed consequentialists to inch closer to common-sense morality, without already buying into agent-relative values. Examples are familiar enough. If consequentialists find torture especially horrendous, their theory can reflect this by placing a high disvalue on it, and if they cherish artistic achievement, they can account for it in their theory of the good by including it as an intrinsic good. Consequentializing, on this picture, is just the latest attempt at modifying consequentialism. As we might remember, Schroeder (2007) himself hesitates to accept consequentializers' newest improvement on value theory, thinking that it leads to problems regarding the *Compelling Idea*, among others.[2] Still, the general idea should be clear enough.[3]

I am not sure there is a similarly continual story to be told about the development in deontology. However, one thing that we have already observed is that deontologists have increasingly found ways to placate

[2] Schroeder (2007, pp. 279 ff.) argues that the Compelling Idea takes its intuitive appeal from the fact that it employs a very simple and ordinary notion of the good: What is good in a way that everyone understands cannot be wrong to be aspired to. But, Schroeder goes on, this notion of good has become increasingly more complicated when consequentialist theories evolved from the likes of Bentham, to Mill, to Moore, and so on. The procedure of consequentializing finally overdoes it. What we end up with, in Schroeder's mind, is very different from the simple notion of the good that we started with and that seemed so attractive.

[3] For a similar story, compare Portmore (2022).

absolutism or have entirely moved away from it. We have seen two tools for doing so in the form of the DDE and the notion of prima facie duties. The DDE allows for some acts that would be forbidden if intended as a goal or means, as long as they can be considered mere side-effects. As Boyle Jr. (1980, pp. 527 ff.) explains, its place in the history of Catholic thinking was to allow some exceptions from an otherwise very rigid moral code. The same goes for the second tool. Prima facie duties were introduced by Ross (1930, pp. 17 ff.) precisely because he could not bring himself to accept that a promise should be kept no matter how horrendous the consequences. So there is some indication of a movement away from the most extreme forms of absolutist deontology in reaction to objections and counter examples.

There is no question that these long-term hypotheses about the developments in consequentialism and deontology would have to be buttressed by historical studies, which I cannot undertake here. There is some more direct evidence, however, that a dialectical tendency played a role in recent developments, especially when it comes to consequentializing. The most outspoken witness for this hypothesis among consequentializers is Portmore. Portmore (2014, p. 209) confesses (his own choice of words!) to having been an adherent of utilitarianism at some earlier stage. However, he recounts, he was not able to reconcile utilitarianism with some of his intuitive convictions about particular cases.[4] As he states very clearly, his main motivation behind consequentializing is to defend consequentialism against the charge of being counterintuitive.[5] Portmore (2022) goes so far as to say that many early philosophers, starting with Mill, should already be considered consequentializers, since they modified their theories in order to come closer to common-sense morality. I think that this way of making the point is too strong. Mill's strategy, as we have seen, differs markedly from consequentializers'. First, the simple recipe that consequentializers employ delivers deontically equivalent counterparts for whole classes of theories, all (plausible) non-consequentialist ones. In contrast, Mill's arguments are only supposed to prove (partial) equivalence for one theory – Kant's – and another – his preferred consequentialist one. Second, consequentializers simply accept any set of verdicts their target theory yields and bend their own theory until it fits the set. Their leverage point is thus the structure of their own theory. Mill makes no use of any such sophisticated machinery,

[4] His example is of having to sacrifice one's own child in order to bring about a state of affairs with a marginally higher net amount of overall goodness.

[5] Compare Portmore (2007, p. 41) and Portmore (2009, p. 331).

like agent-relative values or a constitutive understanding of outcomes. Instead, he argues that Kant's theory would have led to the same verdicts as utilitarianism had Kant been able to free himself from the limitations of his acquired views. Mill's leverage point is thus with Kant's verdicts. Hence, whereas consequentializers match their theory with Kant's verdicts, Mill matches Kant's theory with his verdicts. Mill, and we might say the same for Hare as well, is thus not plausibly considered a proto-consequentializer.[6] That being said, it seems clear that the dialectical development toward more convergence between the different traditions started with Mill (at the latest). At the time consequentializers arrive at the scene, the dialectical pressure might have been heavy enough to lead to full convergence.

This dialectical pressure, we might add, is not just observable over a longer time frame but should feel very familiar to most contemporary ethicists. As one of the first critics of consequentializing puts it:

> Arguing with a consequentialist can be frustrating. Witness a typical sort of exchange: You – a nonconsequentialist, let's assume – begin with your favorite counterexample. You describe some action, a judge's convicting an innocent man to avert a riot, say, or a doctor's murdering a healthy patient for her organs, which, so you claim, would clearly have the best consequences, yet equally clearly would be greatly immoral. So consequentialism is false, you conclude; sometimes a person ought not to do what would have best consequences. "Not so fast," comes the consequentialist's reply. "Your story presupposes a certain account of what makes consequences better or worse, a certain 'theory of the good', as we consequentialists like to say. Consequentialism, however, is not wedded to any such theory. We already knew that combining consequentialism with some theories of the good would have implausible results; that's what utilitarianism has taught us. In order to reconcile consequentialism with the view that this action you've described is wrong, we need only to find an appropriate theory of the good, one according to which the consequences of this action would not be best. You say you're concerned about the guy's rights? No worries; we'll just build that into your theory of the good. Then you can be a consequentialist too." (Brown, 2011, pp. 749–750)

Trying to react to counter examples, consequentialists have been willing to significantly amend their theories. The same, it is plausible to assume, goes for theorists of other traditions. Theories might in this way have converged, making the emergence of deontically equivalent theories more likely.

[6] Peterson (2013, p. 169) agrees.

https://doi.org/10.1017/9781009492454.013 Published online by Cambridge University Press

The Method of Reflective Equilibrium

A second hypothesis is that the converging tendencies have a specific methodological root.[7] That methodological root lies in Rawls's enormously influential introduction of the *reflective equilibrium* account of justification in ethics.[8] According to this methodology, we should try to reach an equilibrium between our preexisting considered beliefs on the one hand and the more theoretical beliefs that try to systematize the preexisting beliefs on the other. Very roughly, this means working back and forth between the two classes of beliefs, revising them in alternating fashion, until we have reached a coherent set of beliefs. As this stage, we have reached reflective equilibrium, implying that the beliefs are more justified than in the beginning.

Reflective equilibrium has its critics, but it is safe to say that it has been incredibly influential in ethical theorizing. This might have increased the tendencies to convergence for a simple reason: If we have to give at least some weight to our preexisting beliefs, we should expect to come up with beliefs less radically counter to common sense than if we do not give any weight to them. This has sometimes been referred to as a *conservative* implication of the method of reflective equilibrium.[9] The point is easily exaggerated. Acceptance of reflective equilibrium does not commit one to conservatism in the sense that one can only ever justify what one believes anyway – coherence and systematization exert a very powerful pressure. Think of a society that tries to bring into reflective equilibrium its theoretical conviction of the equality of men and its standing practice of slavery. Considerations of coherence will likely lead to quite far-reaching modifications of at least one of those convictions. However, it also seems clear that since one part of the procedure is to bring preexisting judgments into accordance with principles, at least some *pro tanto* importance is accorded to those judgments. Compare this to a methodology that, from the start, discards any of our initial judgments and pre-theoretical intuitions and urges us only to follow those verdicts yielded by some stipulated principle. Surely, there should at least be a tendency for philosophers employing the reflective equilibrium methodology to end up with a result that diverges less radically from common sense.

[7] The idea that this is what is behind some of the recent developments in normative ethics was pointed out to me in conversation by Brad Hooker.

[8] Compare Rawls (1971) for the classical depiction of the method. Compare also Rawls (1951) for a much earlier, slightly different outline of the idea.

[9] For discussions see Copp (1985, pp. 144–146) and Scanlon (2003, pp. 149–151).

Is there any evidence that adherence to reflective equilibrium might have played a role in the recent developments in normative ethics? At least when it comes to Parfit, this seems to be the case. He approvingly quotes Scanlon who, commenting on the method of reflective equilibrium, writes:

> [...] this method, properly understood, is [...] best way of making up one's mind about moral matters [...]. Indeed, it is the only defensible method: apparent alternatives to it are illusory. (Parfit, 2011a, p. 367)[10]

Hooker (2020, p. 5) accordingly surmises that a presumption for reflective equilibrium methodology is *woven* into *On What Matters*.[11] De Lazari-Radek and Singer agree. They think that acceptance of the method of reflective equilibrium might have played a role in Parfit's thinking leading up to his Convergence Argument:

> [...] like so many contemporary moral philosophers, he accepts the model of reflective equilibrium made popular by John Rawls, and this leads him to be reluctant to reject too many of our common moral judgements. (de Lazari-Radek and Singer, 2016, p. 294)[12]

Hence, there is some direct as well as some indirect evidence that adherence to reflective equilibrium might also have something to do with the development of deontically equivalent theories, at least in Parfit's case.[13]

Interestingly, Carrier (2011, pp. 190–191) argues that the rise of underdetermination in science also came on the back of a methodological reorientation. In his view, it was the switch from the inductivist model to the hypothetico-deductivist model that brought underdetermination into focus. As the standards for what counts as an acceptable way of coming up with scientific hypotheses were relaxed, allowing not only hypotheses that are directly suggested by the evidence, it became increasingly more likely that different theorists would come up with alternative theories. As Carrier (2011, p. 191) puts it: "Underdetermination is an unintended by-product of the methodological transition from inductivism to hypothetico-deductivism." I am, to be clear, not claiming that the introduction of the method of reflective equilibrium has had the same influence on the development of moral underdetermination as the introduction of

[10] Compare also Parfit (2011b, p. 544).
[11] Citing several other passages where reflective equilibrium plays a role.
[12] Recall also how Darwall (2014, pp. 80–81) observes a turn in Parfit's thinking from a more radical to a more conservative outlook.
[13] There is also one consequentializer who can be linked to the method of reflective equilibrium. Portmore (2011, pp. 112 ff.) makes prominent use of the method when it comes to his explanation of why consequentialists can come to the same verdicts as non-consequentialists without their theory becoming a fifth wheel.

hypothetico-deductivism did in science. However, the idea that methodological developments were influential in bringing about the situation of underdetermination in both science and ethics might have something to be said for itself.

Rendering the Advent of Underdetermination Less Suspicious

If the above considerations are correct, a combination of repeated adjustments to meet objections and a methodology that favors less radical stances might have had a significant influence on the situation as it presents itself in normative ethics at this time, with theories converging on their verdicts.[14] This would certainly be interesting from a merely historical perspective. However, if true, it also helps alleviate two kinds of doubt that one might have concerning moral underdetermination.

First, some might find the relatively recent emergence of moral underdetermination a disappointing sign. Sure, the objection could run, we can come up with ever more convoluted theories, which at one stage will agree on verdicts. But these versions have nothing to do with what the traditions originally stood for. Underdetermination might thus fail to be interesting, exactly because it is the result of these dialectical and methodological steps being taken too far. I understand where this disappointment comes from. The theories we have considered have indeed come a long way from Kant and Bentham. The question then becomes how far we can take the modifications without losing what is distinctive about the moral traditions. Sinnott-Armstrong, regarding consequentialism, holds that:

> In actual usage, the term 'consequentialism' seems to be used as a family resemblance term to refer to any descendant of classic utilitarianism that remains close enough to its ancestor in the important respects. Of course, different philosophers see different respects as the important ones. Hence, there is no agreement on which theories count as consequentialist under this definition. (Sinnott-Armstrong, 2015, pp. 4–5)[15]

What is close enough? I have proposed that a theory needs to claim that acts are right if and only if *and because* they lead to the best outcomes in order to qualify as consequentialist. Yet that only transforms the question to: What counts as an outcome and what counts as best? Opinions will diverge again. Something similar is probably true for the kinds of

[14] Notice that this is not the same issue as whether cases of existing underdetermination are *permanent* or *transitive*. Moral underdetermination might have only come into play at an advanced stage of the theoretical debate, but the cases of underdetermination that develop at that stage can nevertheless prove permanent.

[15] The passage is quoted in Portmore (2009, p. 335).

non-consequentialist theories deontologizers come up with, and we have also encountered philosophers who do not consider Parfit's preferred theories as representative of the main traditions.

My view, in contrast, is that we should go along with and even welcome the new developments. There is something strangely traditionalist in charging ethicists for wanting to improve their traditions. We have to take seriously that a majority of theorists are no longer satisfied with the original theories of Kant and Bentham and have moved on. In this context, it is apt to consider again Parfit's memorable words:

> After learning from the works of great philosophers, we should try to make some more progress. By standing on the shoulders of giants, we may be able to see further than they could. (Parfit, 2011a, p. 300)

Additionally, as I have pointed out in Chapter 5, in contrast to the algorithms in science, consequentializing and deontologizing have not been introduced with the sole purpose of proving underdetermination, and neither has Parfit's Convergence Argument. Instead, if the above reasoning is apt, the projects follow independent dialectical and methodological logics. Granted, underdetermination in ethics might be a relatively recent phenomenon that only came into prominence at a later dialectical and methodological stage. This is not to belittle it, though. Instead, underdetermination in ethics should be taken seriously precisely because it has become increasingly relevant in recent times.

There is a second suspicion that is likely to have arisen in some readers' minds at this stage. The suspicion goes like this. Assuming that the different traditions do indeed give mutually exclusive explanations for the rightness or wrongness of acts, why then would they come to the same conclusions anyway? If there is such a deep theoretical divide, one would expect that this divide manifests itself in extensional differences. Since it does not, we might begin to question the alleged explanatory differences. On the underdetermination view, it might seem merely contingent that a theory that is based on the evaluation of consequences and one that cherishes duties or rights would arrive at the same deontic conclusions. Is that not too much of a coincidence? And is it not in that case more plausible to assume that the explanatory differences are not radical after all?

Having some understanding of the dialectical and methodological developments that lead to underdetermination helps alleviate these doubts. The different traditions have made considerable dialectical moves toward each other, and a specific methodology preferred by many ethicists today further strengthens agreement. But that does not mean that ethicists have given

up the explanatory claims that distinguished them from the beginning. If the two hypotheses are correct, it is no mere coincidence that the rival theories would converge in their verdicts; rather, it is the result of independently motivated dialectical and methodological developments from within normative ethics. This should help alleviate the suspicion of convergence being too much of a coincidence.

8.2 The Puzzle Solved?

What got this book started was the observation about two seemingly irreconcilable tendencies in contemporary normative ethics. On the one hand, the textbooks pitch the different traditions of moral theorizing against each other, extensionally and explanatorily. On the other hand, several authors have recently argued that a wide convergence between the rival traditions is possible. Over the course of the preceding chapters, I have outlined a new view that might go some way toward explaining the two tendencies. The underdetermination view upholds the second tenet of the Textbook View: The different traditions are indeed incompatible when it comes to their fundamental explanatory claims. In contrast, the underdetermination view contradicts the first tenet: Theories from different traditions need not disagree regarding the verdicts they yield for particular cases. Both the Textbook View and proponents of the newer projects might thus be partly correct, although they only capture one side of the phenomenon of moral underdetermination.

Two features of the new view make it preferable to other interpretations of the recent projects. The first one has popped up throughout the book, especially with regard to what I consider to be the main rival interpretation, the notational variants view. The underdetermination interpretation invites us to think of the problem that extensionally equivalent moral theories pose in epistemological, not semantic, terms. This, I have argued, allows us to uphold a more plausible moral semantics. This advantage has a parallel in the philosophy of science. One of the main advantages of van Fraassen's view in science is that is breaks with the implausible semantic views of logical positivists. If we wanted to put the point provocatively, we could say that the underdetermination view allows us to similarly break with the moral analog of a positivist undercurrent in some of the other interpretations of the recent projects. If we are skeptical about moral metaphysics, we need not claim that the seemingly metaphysical claims about moral explanation can somehow be reduced to claims about action-guidance. Instead, the underdetermination view raises doubts that anyone

really knows what makes an act right or wrong, and constructive deonticism might reconcile us with the idea that we don't really need to know anyway.

The second advantageous feature of the underdetermination view is that it retains at least some of the adversarial spirit of the textbooks, by upholding the second tenet of the Textbook View. This might seem a strange advantage. Why should it be positive that the underdetermination view does not fully reconcile the rival traditions? Yet the reason is not difficult to understand. The underdetermination view allows us to at least partly make sense of the widespread understanding of the moral traditions as antagonists. Most of the other views we have encountered make it very difficult to understand how and why the traditions are commonly assumed to be disagreeing. Morgan makes this point about Parfit:

> [...] for the bulk of the last century the territory of moral theory was widely held to be divided between consequentialists and deontologists, and the positions to be fundamentally opposed to one another. So if Parfit is right, the question does arise: how could so many people have been so confused? (Morgan, 2009, p. 42)

Parfit's conciliatory view makes it difficult to explain the widespread impression of antagonism between the traditions. So too do some of the consequentializers' views. How could it be that we have thought of some theories as genuinely different from consequentialism when they are in fact part of that family? These other interpretations thus have great difficulties explaining why it is that most of us have been, and are still, thinking of the moral traditions as genuinely in opposition to each other.

The underdetermination view has a simple explanation for this: People are right. Granted, the underdetermination view then still owes us an explanation for why most philosophers have also subscribed to the first tenet of the Textbook View. However, I think that this is a more manageable task since the second tenet is much more deeply entrenched in our understanding of moral theories than the first. This can be brought out by the dialectics of a very familiar piece of moral reasoning. Recall the depiction by Brown (2011, pp. 749–750) of a typical discussion between a consequentialist and a non-consequentialist. If asked what their favorite counter example to consequentialism is, many non-consequentialists might initially point to some specific judgment, like the harvesting of organs from one person for the benefit of others, or some other gruesome verdict that they presume will get the consequentialist into trouble. Yet, whatever the verdict is, consequentialists can always counter that they do not approve

of that judgment, for it does not match their theory of the good. This is what Brown finds so frustrating. However, the quarrel typically does not end at this point. Instead, non-consequentialists will next retort that their opponents might well come to the right conclusion here, but for the wrong reason. That is, their opponents might be able to make a place for the correct verdicts, but their explanation is nevertheless inaccurate. In other words, non-consequentialists will fall back on some kind of disagreement that they take to be more fundamental than the mere extensional one, which is a disagreement at the theoretical level.

The topic of underdetermination is a prime example of Putnam's claim, cited in the Introduction, that the problems and arguments in different fields of philosophy resemble each other. In this book, I have argued that there is a highly instructive structural analogy between science and ethics: Just as scientific theories can be underdetermined by the empirical evidence, so moral theories can be underdetermined by our considered judgments or intuitions about particular cases. Ethicists might learn a lot from this analogy. However, this is not meant to be disparaging toward ethicists. I think that most of the insights that can be gained are not about actual flaws in ethical theorizing that the philosophy of science would teach us to correct. Instead, the lessons are about developments that have already taken place but which have not been appreciated in full generality or have not been given a systematic explanation. The philosophy of science is helpful in this regard for the simple reason that people have long been discussing an analogous phenomenon to what some of the newer strands in moral theorizing are leading to. Before we can assess with certainty whether the underdetermination view ultimately proves to be the correct explanation, perhaps as much time will be needed in ethics as well. However, I hope to have at least succeeded in presenting it as a view worthy of further investigation.