# *Introduction*

## I.1 A Puzzle

In normative ethics, a relatively small number of traditions of moral theories, such as Kantianism, consequentialism, or contractualism, take center stage. Students are introduced to these traditions early on in their studies, while normative ethicists, in much of their day-to-day work, either employ one tradition's framework or argue directly for or against (a version of) a tradition. These few traditions, it is no exaggeration to say, are many ethicists' bread and butter.[1] Considering this fact, it might come as something of a surprise that ethicists have highly discordant views about how these traditions stand with regard to each other. More specifically, one might be puzzled by two seemingly irreconcilable ways of looking at the landscape of moral theorizing.

On the one hand, the standard view of the relation between the major traditions is one of antagonism. Conventional wisdom has it that Kantianism, consequentialism, and contractualism stand in stark contrast. They represent rival strands of moral thinking that have developed into elaborate, irreconcilable theoretical systems. More precisely, the textbooks tell us that the traditions of moral theorizing exhibit two levels of disagreement. First, the traditions disagree on the *particular deontic verdicts* they yield, that is, which specific acts they hold to be right or wrong, forbidden, obligatory, or allowed.[2] We can refer to this as the *extension* of moral theories. Second, the traditions disagree when it comes to the question of what *makes* particular

[1] Of course, this is not to say that all that ethicists ever do is engage with these traditions. In some areas of both normative and applied ethics, discussion can take place without explicit reference to these traditions. Still, it should be uncontroversial to note that the main traditions play a large overall role. Compare Shafer-Landau (2012, pp. 14–15) for this point.

[2] In order not to unduly disturb the reader's flow, I will often use the phrases *right or wrong* or *ought* when I want to refer to all the deontic verdicts a theory yields. Unless otherwise specified, this phrase should be understood as a shorthand for *right or wrong, mandatory, allowed, disallowed, and so on*. However, it does not include aesthetic verdicts or those of etiquette.

acts right or wrong, that is, the *explanations* they put forward for why some acts are right or wrong. Kantians will refer to ideas like autonomy, the good will, or humanity as an end in itself. Consequentialists, by contrast, emphasize the notions of an (actual or foreseeable) outcome, the overall utility of an act, or its ranking on a scale of goodness. Contractualists, finally, use the ideas of a (hypothetical) agreement, some kind of initial position, or cooperation between rational (self-interested) agents.

We can call what I have just described the *Textbook View* of the moral traditions.[3] It consists of two tenets:

> (I) *Tenet of Extensional Disagreement*: The moral traditions arrive at different particular deontic verdicts, that is, verdicts about which acts are right or wrong, permissible, obligatory, or forbidden.
>
> (II) *Tenet of Explanatory Disagreement*: The moral traditions give competing explanations of what makes an act possess its deontic status.[4]

Of course, this is not all that the textbooks have to tell us about moral theories, and there are many exceptions, as is to be expected in philosophy. However, I think it is fair to say that the Textbook View encodes a very widespread way of thinking about the relation between the moral traditions. Even though many ethicists might not consciously subscribe to this view, it should fit with a majority's understanding of their own discipline.

On the other hand, for a long time there have been dissenting voices arguing that considerably more agreement is possible between the traditions. A number of philosophers have recently taken this idea to a new level. First, Derek Parfit, in his monumental 2011 book *On What Matters*, has argued that the best versions of three of the most prominent traditions of moral theorizing actually agree about all that matters, that is, they accept the same set of principles about what is right or wrong. Second, a group of philosophers have argued that they can *consequentialize* any (plausible) rival theory. By this they mean that for any (plausible) non-consequentialist theory, they can come up with a consequentialist counterpart, which is deontically equivalent, that is, leads to the same set of deontic verdicts. This has brought yet another group to the scene, whom we can call *deontologizers*. They claim that they can do for non-consequentialism what

[3] An especially concise example is Tännsjö (2002). We will encounter more in Chapter 2.

[4] The idea of the Textbook View with its two tenets takes its inspiration from Dougherty (2013, pp. 527–530). However, I prefer to describe the second tenet in terms of *explanation* instead of *intension*. I also do not restrict disagreements concerning the second tenet to those having to do with the *agent-neutral* versus *agent-relative* distinction.

consequentializers promise for consequentialism, namely provide deontically equivalent non-consequentialist theories for every consequentialist theory. These projects challenge the Textbook View. They directly challenge the first tenet since, if they are right, the rival traditions need not disagree about the verdicts they yield. In addition, most proponents of these projects also challenge the second tenet. They put forward interpretations that range from the claim that we are all consequentialists (or deontologists), to the claim that the moral traditions are merely notationally different, to the claim that the traditions can be reconciled or even combined.

What we thus see are two factions of ethicists. One faction, the majority, thinks of the moral traditions as mutually exclusive frameworks. Another faction, a defiant minority, sees no major differences. How, we might ask, is this possible? How can people have such diverging assessments of the most basic theoretical dynamics in their primary domain of study? This book can be read as an attempt to shed light on this question from a new perspective. Perhaps both sides are right, but only partly so. Moral theories might sometimes agree on all the verdicts they yield about what is right or wrong. At the same time, they might nevertheless advance incompatible explanations for why these acts are right or wrong. This would mean that we are facing a so-far neglected phenomenon: the *underdetermination of moral theories*.

## I.2 Underdetermination: From Science to Ethics

The inspiration for this idea comes from the philosophy of science. Following Pierre Duhem (1906) and W. V. O. Quine (1951), philosophers of science have discussed a phenomenon that looks remarkably similar to the situation described above: the *underdetermination of scientific theory by evidence*. Sometimes, it is held, all the evidence available to scientists is insufficient to decide between two or more theories that make radically different claims about the hidden makeup of the world. The choice between such theories is left open by the evidence; the theories are *underdetermined*.[5]

The underlying hypothesis of this book is that there is a *structural analogy* to be found here: Just as scientific theories can be underdetermined by

[5] I will sometimes speak of *theories being underdetermined* and sometimes of *the choice between theories being underdetermined*. There is no difference in meaning intended here. The first phrase is just a short form of the second.

the empirical evidence, so moral theories can be underdetermined by our considered judgments or intuitions about particular cases.[6] This analogy has not been investigated in detail. Of course, ethicists have not overlooked the fact that moral theories from different traditions might arrive at the same verdicts while remaining explanatorily incompatible. Indeed, this very point has been raised with regard to the projects just mentioned.[7] However, as far as I know, the first explicit mention of this specific analogy to science is due to Franz Dietrich and Christian List. In the process of outlining a new way to formally represent moral theories, they point out that their *reason-based representation* of moral theories can help to:

> [...] shed light on an important but still underappreciated phenomenon: different moral theories may coincide in all their action-guiding recommendations, despite arriving at them in different ways [...]. Put differently, the same action-guiding recommendations may be explained in more than one way. For example, some deontologists and some consequentialists may agree on all 'ought' statements but offer different explanations for them. Our reason-based approach allows us to formalize and investigate the generality of this phenomenon: the *underdetermination of moral theory by deontic content*. (Dietrich and List, 2017, p. 422)

Dietrich and List mention the possibility of moral underdetermination with regard to both Parfit's project and the consequentializing project. I have subsequently developed the analogy in a series of papers in a less formal way.[8] However, due to their relative brevity, these contributions inevitably only shed light on some aspects of the topic.

The goal of this book is to investigate the analogy in its full generality, thereby assessing its upshots for both normative ethics and metaethics. For this purpose, I will not be defending a *specific* underdetermination thesis for the moral realm. This is not for lack of confidence in the illuminating power of the analogy. Instead, as the philosophy of science

[6] As I shall explain in more detail in Chapter 2, the analogy is a *structural* one because it does not presuppose that there is a similarity in the nature of the two domains – ethics and science – but only one with regard to a specific structure of theory choice.

[7] Compare Suikkanen (2014, p. 104) and Portmore (2011, p. 109).

[8] With regard to Parfit (Baumann (2018, 2021a)), the consequentializing project (Baumann (2019)), as well as possible metaethical upshots (Baumann (2021b, 2022)). Some of the chapters in the present book draw on these papers, especially Chapters 3, 4, and 6. I should note that I came up with the analogy independently of Dietrich and List, having worked on the topic as early as 2013. However, their insights have greatly furthered my understanding.

shows, underdetermination is a multifaceted phenomenon that can take different shapes. To quote Larry Laudan:

> More or less everyone [...] agrees that 'theories are underdetermined' in some sense or other; but the seeming agreement about that formula disguises a dangerously wide variety of different meanings. (Laudan, 1990, p. 269)

This, we shall see, is true for ethics, too. Moral underdetermination can come in different shapes. Defending just one specific thesis would thus be stifling. Instead, the best way to describe the current book's aim is that it seeks to provide the first sustained analysis of the phenomenon of moral underdetermination. It is an investigation into how the debate about scientific underdetermination can help inform our understanding of some recent developments in normative ethics, as well as an attempt to give new impulses to our general understanding of the realm of the moral – normative ethics and metaethics.

Attempting such a venture poses some unique challenges. The idea of moral underdetermination is at odds with the Textbook View, which holds that theories from different moral traditions come to different deontic conclusions and are therefore not underdetermined. Building a case against the Textbook View from scratch would be a herculean task. Fortunately, I have allies in the proponents of the three aforementioned projects – consequentializing, deontologizing, and Parfit's *On What Matters* – for they argue precisely that theories from different traditions can agree on their extension. Less fortunately, these authors have their own views about what their projects amount to, none of which is compatible with an interpretation in terms of underdetermination. I cannot therefore simply rely on the results of these projects. Instead, the quest is twofold. On the one hand, I need to show how these projects challenge the textbooks. On the other hand, I need to argue that *pace* what proponents of these projects may think, their results are best understood in terms of underdetermination.

At least to some degree, then, I will have to rely on what proponents of the three projects profess to have established. Although I will offer some general assessments of plausibility, all three projects involve quite complex reasoning in normative ethics that I cannot assess in detail. However, in this regard, I feel I am no worse off than philosophers of science are when they try to assess the phenomenon of underdetermination in science: The specific examples of underdetermination they investigate are from the special sciences, and we cannot expect them to assess the science behind all those examples in detail themselves. To some degree, they have to rely on what scientists report. However, that does not threaten their philosophical insights; it merely makes some of them conditional on matters that the

philosophers cannot decide themselves. In the same way, I sometimes have to rely on preliminary results that normative ethicists report, but I believe that the philosophical insights are not tainted by this.

The philosophy of science might not be the most obvious place to look for analogies with the field of ethics. However, I agree with Hilary Putnam that:

> [...] the unfortunate division of contemporary philosophy into separate 'fields' (ethics, epistemology, philosophy of mind, philosophy of science, philosophy of language, philosophy of logic, philosophy of mathematics, and still others) often conceals the way in which the very same arguments and issues arise in field after field. (Putnam, 2004, p. 1)

It is in this spirit that my book makes the case that we should study the underdetermination debate in the philosophy of science in order to gain insights for ethics. For the most part, this means informing our understanding of *normative* ethics. It is normative moral theories that (might) turn out to be underdetermined; it is a view of normative ethics, the Textbook View, that is challenged; and many of the insights that are generated by transferring the idea of underdetermination from science to ethics concern normative ethics. Since I take these results to be largely independent of metaethical assumptions, I try to remain as metaethically neutral as possible in Parts I and II. That being said, I do take moral underdetermination to have interesting repercussions for metaethics as well. These find their full expression in Part III.

## I.3 Demarcating the Main Idea

Before I lay out the plan for the book, I want to say a bit more about how I understand its main idea. Since giving a positive account of this will occupy the whole of Chapter 2, a negative characterization will have to suffice for now. The term *underdetermination* has been applied to issues in ethics before, but not to draw the same analogy to science. So here is what the book is *not* about.

First, the idea is not that moral theories are underdetermined by non-moral, empirical data. This is a claim that has some currency in metaethics. For example, Owen Flanagan, commenting on Quine's views on ethics, holds that:

> Morals are radically underdetermined by the merely descriptive, the observational; but so too, of course, are science and normative epistemology. All three are domains of inquiry where ampliative generalizations and underdetermined norms abound. (Flanagan, 2006, p. 443)

Moral underdetermination, in Flanagan's sense, means that all our descriptive, non-normative observations cannot determine what the correct moral theory is. This is not the kind of underdetermination that I have in mind. What interests me is whether genuinely moral commitments or intuitions (about what is right and wrong, forbidden, required, or allowed) underdetermine genuinely moral theories. This is a wholly different question. Indeed, to avoid confusion with this alternative understanding of moral underdetermination, unless stated otherwise, I will assume that there are *no* non-moral disagreements that underlie the relevant disagreements between the rival traditions. This is certainly a simplification. However, it serves to bring out the idea of moral underdetermination, as I understand it, more clearly.

Second, the idea is not that moral theories or principles leave us undecided when it comes to concrete moral decisions. Moral theories, one might think, are simply too general and unspecific to provide us with directions that are sufficiently determinate for us to know what to do in a particular case. This is another way in which one could quite naturally understand the term moral underdetermination. However, moral underdetermination, as I understand it, is the exact mirror image of this idea. It does not concern whether our judgments about what we should do are underdetermined by our theories; rather, it is about the opposite determination relation: the way in which our particular considered judgments or intuitions may, or may not, determine which theory is correct.

Third, the idea is not that our considered judgments and intuitions underdetermine theory choice because too few of them are shared or they are too vague or there is indeterminacy about what the correct actions are.[9] This idea is at least in the same ballpark as the one I want to discuss, since it does have the right direction of fit; that is, particular judgments or intuitions underdetermine theory choice. However, moral underdetermination, as I understand it, does not rely on any such defect with regard to the evidence. It holds that theories are underdetermined even if, or rather because, there is agreement about what the correct judgments are in particular cases. Indeed, at least some versions of the underdetermination thesis that I will discuss explicitly state that theory choice is underdetermined by all possible moral judgments. We can acknowledge, in accordance with my understanding, that the evidence would in principle be sufficient to determine which theory is correct, were it not for the fact that there is some other theory that can account equally well for all the evidence.

[9] For this last point, compare the contributions summarized in Elson (2016).

In sum, the idea is neither that the moral is underdetermined by the non-moral, nor that moral theories leave our particular choices underdetermined, nor that theories are underdetermined because the evidence is somehow lacking. Instead, theory choice is underdetermined by our considered judgments or intuitions because mutually exclusive moral theories can account equally well for all of them. At least this is the idea that I will investigate in the course of the book.

## I.4 The Plan for the Book

The book is structured into three Parts. Part I sets the stage by introducing the discussion of underdetermination in the philosophy of science and outlining the analogy to ethics.

Chapter 1 covers the background in the philosophy of science. The goal here is twofold. First, I aim to provide the reader with a firmer grasp of the general idea of scientific underdetermination. For this purpose, I start with some examples to illustrate the phenomenon before I introduce the two main progenitors of the idea, Pierre Duhem and W. V. O. Quine. Second, I provide a more systematic outline that can inform the later discussion in ethics. I introduce the main argumentative strategies that have been employed to argue for underdetermination in science. I then make a number of distinctions between different versions of underdetermination to provide a picture of the various forms that underdetermination can take.[10]

Chapter 2 draws out the structural analogy between science and ethics: the idea that just as scientific theories can be underdetermined by the empirical data, so moral theories can be underdetermined by our considered judgments or intuitions. The first section is devoted to the notion of a moral theory. I introduce what I take to be an independently plausible view that ascribes two functions to moral theories: accounting for our considered judgments or intuitions and explaining them. I then go into the details of both functions. The second section is about the analogs of the evidence of scientific theories, which I take to be considered judgments or intuitions about particular cases. The third section considers two possible problems with the analogy.

Having established the background, Part II investigates the phenomenon of underdetermination as it relates to normative ethics.

[10] I do not discuss objections at this stage so as not to render the book too top-heavy. Instead, these will be taken up when similar problems arise in the ethical case.

Chapter 3 is concerned with what, I argue, should be viewed as the most in-depth case study of moral underdetermination: Derek Parfit's argument in *On What Matters* to the effect that the best versions of three of the most famous moral traditions arrive at the same verdicts about what matters. Since Parfit himself does not think of his project in terms of underdetermination, my argument here requires a two-step approach. First, I outline how Parfit challenges the Textbook View, focusing on what he calls the *Convergence Argument*. Second, I argue that contra Parfit's own understanding, the best way to think of the results of the Convergence Argument is in terms of moral underdetermination. If my argument proves successful, we are thus presented with one very detailed and very prominent case study of moral underdetermination. Still, one case study may not be thought enough to put the phenomenon of moral underdetermination firmly on the map.

Chapter 4 thus changes gear. I look at two projects that aim to establish far more radical conclusions, so-called consequentializing and deontologizing. Proponents of these projects try to come up with a simple mechanism to produce deontically equivalent counterparts to any consequentialist or non-consequentialist theory. I examine both projects in turn, explaining how they work on a technical level as well as how and why we should interpret their results, once more *pace* their proponents, in terms of underdetermination. Interestingly enough, this discussion mirrors one in the philosophy of science, where some defenders of underdetermination have tried to establish more pervasive forms of underdetermination by generating algorithms that produce empirically equivalent theories for any existing scientific theory. If consequentializing and deontologizing can be interpreted in the same way, then we are presented with a much more far-reaching version of moral underdetermination.

Chapter 5 takes stock and considers the bigger picture. I first make explicit the analogy to the three strategies outlined in Chapter 1 and what we can learn from this about their relative prospects and plausibility. I next draw a map of underdetermination in ethics, assessing each version of the thesis that has been identified for the scientific realm in Chapter 1 in terms of its status in ethics. To finish, I outline what I take to be the three most important lessons of moral underdetermination for the future of normative moral theorizing.

On this note, Part III leaves normative ethics behind, turning to metaethics.

Chapter 6 explains how the results in normative ethics might give rise to a skeptical challenge. I start by revisiting Duhem and Quine and how

they thought underdetermination should impact our view of science. I then outline why I understand the major challenge that moral underdetermination poses to be an epistemological, not a semantic, one. This leads me to sketch out the general lines of a skeptical argument, building on a similar argument in the philosophy of science. In conclusion, I address two possible objections, arguing that the skeptical argument in ethics is at least as plausible as, if not more plausible than, its counterpart in the philosophy of science.

What might the success of the skeptical argument entail? Chapter 7 provides one possible answer. Perhaps the correct reaction to the skeptical argument is to accept a novel metaethical position: *constructive deonticism*. This position is structurally analogous to one of the most discussed antirealist positions in science, Bas van Fraassen's *constructive empiricism*. I thus start with some background on van Fraassen's view. I then introduce and flesh out the new position in ethics. Following this, I discuss how constructive deonticism should be classified among other metaethical positions, arguing that this is less than clear and that it might actually prompt us to rethink how the metaethical realism debate has so far been framed. Finally, I point out what I consider to be the two most important challenges that a defender of constructive deonticism would have to answer in the future.

Chapter 8 brings together some of the different strands and concludes the book. I consider two hypotheses about how we ended up with underdetermination in ethics, one citing dialectical advancements, the other pointing to the method of reflective equilibrium. If correct, these hypotheses help alleviate some doubts about the fact that the analogy to science has only become relevant quite recently. Finally, I return to the puzzle and explain why I take the underdetermination view to offer the best answer to it.