

Methodology

2.1 Introduction

This book is a work of normative theory – specifically, an attempt to provide a general framework that can illuminate and address ethical issues that arise in biomedical contexts. Such normative theorizing requires an appropriate methodology. The methods of the natural sciences – involving the collection of data, the testing of hypotheses through experimentation, and so forth – are insufficient for drawing conclusions about what ethically ought to be done. In this chapter, we describe and defend the methodology that we use in the rest of the book.¹

Our reasons for explicitly laying out our methodology are threefold. First, doing so facilitates critical engagement. For someone who disagrees with us, it can be very helpful to identify whether the disagreement is a matter of starting from a different set of values, reasoning in different ways from similar starting points, or failing to meet some shared standards for ethical reasoning. Second, doing so displays the standards to which we think we should be held. The arguments we make over the course of the book should be explicable and defensible in terms of the methodology we describe here. Third, presenting our method may be helpful to readers who do not work in philosophical ethics. It is common for nonphilosophers to question how normative theorizing is done: “Do you collect data on what people think is ethical? If not, aren’t you just stating your opinions?” Explaining the methodology and demonstrating how to use it may provide both a justification for how we proceed and some helpful examples for readers who are new to bioethics or ethical theory.

¹ Each of us has described his methodology in ethics elsewhere. See David DeGrazia, *Taking Animals Seriously* (Cambridge: Cambridge University Press, 1996), chap. 2; and Joseph Millum, *The Moral Foundations of Parenthood* (New York: Oxford University Press, 2017), 4–14. Our discussion here is consistent with those accounts while going beyond them in various respects.

Although normative theorizing and empirical research use different methods, empirical data are highly relevant to normative work in bioethics.² It is generally impossible to make action-guiding ethical recommendations in a particular case without taking into account the empirical facts that characterize that case. For example, in thinking about whether to follow a family's request to discontinue treatment for a terminally ill patient, we need to ascertain such matters as the following: what the prognosis would be with and without different treatments, whether the patient is suffering, what the options are for palliation, whether the patient's wishes are known, and what resources are available to the hospital. Answering these questions requires empirical information about the case (such as what wishes the patient expressed) and inferences drawn from empirical studies (such as the likelihoods of different possible outcomes of a treatment).

We begin this chapter with a description of our methodology, which we take to be a version of the method of *reflective equilibrium* that is widely used in philosophy and bioethics. We then describe some methodologies for normative theorizing in bioethics that are often advanced as alternatives to this methodology. Some, such as casuistry, we think are better understood as versions of the method of reflective equilibrium, at least as they are typically practiced. Others, such as particularism and foundationalism, we reject. Next, we defend the method of reflective equilibrium against some prominent criticisms, including skeptical reactions about the use of intuitions that recent work in experimental philosophy has engendered in some commentators. Finally, we turn to metaethics and clarify what we are, and are not, assuming about the nature and foundation of ethics.

2.2 The Method of Reflective Equilibrium

The basic idea behind the method of reflective equilibrium is relatively simple. We start with our existing ethical beliefs about cases and principles, weed out those that are thought to be unreliable, and then adjust the remaining set in order to make it as coherent as possible.³ The final goal – which may never be

² There is a substantial amount of rigorous and valuable empirical bioethics research, and multiple academic disciplines contribute to bioethics scholarship. For an overview of the disciplines and their methodologies, see Jeremy Sugarman and Daniel Sulmasy, *Methods in Medical Ethics* (Washington, DC: Georgetown University Press, 2010).

³ We use the language of "beliefs" because we find it natural to speak of ethical beliefs. We do not intend this usage to beg any questions regarding whether ethical judgments can be true or aim to express truths (see Section 2.5).

reached but stands as a regulative ideal – is a set of principles that fit together as a single theory and which, along with the relevant empirical facts, entail the moral judgments about cases that we think are correct. In the following paragraphs we fill out this idea and make it more precise.

Terminology and Scope

A couple of points regarding terminology and scope merit mention at the outset. First, some writers use “principles” to pick out only the most general of ethical judgments. For example, Beauchamp and Childress distinguish “principles” from “rules” such that principles are more “general and comprehensive” and rules are “more specific in content and more restricted in scope.”⁴ We make no such distinction. We use “principle” to refer to a universal normative statement, no matter how general or specific. Occasionally, we follow common usage in speaking of “rules” or “rules of thumb,” but do not mean these as technical terms.

Second, our initial set of ethical beliefs are not just those judgments that we have already explicitly made. Implicit ethical beliefs can be elicited. The use of cases to prompt intuitions is a common way to demonstrate to someone that they have beliefs of which they were not aware. For example, someone might be persuaded that they already believe in an ethical difference between killing and letting die when they discover that their reaction to a case in which a physician can administer a lethal injection to a patient in terrible pain is different from their reaction to a case in which a physician can withdraw life-support measures from a similar patient. Notice, too, that these ethical beliefs come in different forms. We have intuitive reactions to particular actions or cases – “It would be wrong for Dr. Gomez to kill her patient.” We also have intuitive reactions about the plausibility of ethical principles – “It is wrong to provide more benefits to one person than another solely on the basis of gender.”

Third, though the ultimate goal of reflective equilibrium is a set of principles, a key part of the process involves articulating what those principles *mean*. Often there are terms used in candidate ethical principles whose meaning is unclear or disputed. Such terms include “well-being,” “harm,” “autonomy,” “equality,” “voluntariness,” and so on. Settling on the correct principles must then include settling on the correct understanding of these terms. For example, the principle of nonmaleficence is a

⁴ Tom Beauchamp and James Childress, *Principles of Biomedical Ethics*, 7th ed. (New York: Oxford University Press, 2013), 14.

prohibition on causing harm. Assuming that we start with the considered judgment that some version of this principle is correct, the process of reflective equilibrium will involve working out the conditions under which it applies (e.g., is it wrong to harm someone who gives consent, or to harm one person to prevent harm to another?). But we cannot apply the principle without also knowing what harm consists in. As we discuss in Chapter 4, there are different accounts of harm. These different accounts can themselves be assessed and amended on the basis of their fit with our considered judgments about principles and cases.

Finally, for clarity of exposition, we mostly restrict our explanation of the method of reflective equilibrium to ethical judgments about principles and cases. However, this does not exhaust the relevant considerations that may be used in moral argument. In the process of attaining what Norman Daniels described as “wide reflective equilibrium” we may bring up all sorts of beliefs about values, reasons, and metaphysics. Daniels writes:

Though we may be committed to some views quite firmly, no beliefs are beyond revision. . . . I include here our beliefs about particular cases; about rules and principles and virtues and how to apply or act on them; about the right-making properties of actions, policies, and institutions; about the conflict between consequentialist and deontological views; about partiality and impartiality and the moral point of view; about motivation, moral development, strains of moral commitment, and the limits of ethics; about the nature of persons; about the role or function of ethics in our lives; about the implications of game theory, decision theory, and accounts of rationality for morality; about the ways we should reply to moral skepticism and moral disagreement; and about moral justification itself.⁵

In the arguments of later chapters concerning personal identity, procreation, and moral status, this breadth of relevant considerations should become clear.

From Initial Beliefs to Considered Judgments

From the set of initial beliefs about cases and moral principles we select just those that we think have sufficient credibility. These are the *considered judgments* that form the data for our ethical theory. Our initial ethical judgments may be eliminated as candidates for considered judgments for various reasons. One is that we lack confidence in those judgments; that is,

⁵ Norman Daniels, “Wide Reflective Equilibrium in Practice,” in Daniels, *Justice and Justification* (New York: Cambridge University Press, 1996), 333–352, at 338–339.

we are uncertain about whether they are correct.⁶ Cases in which we are uncertain about whether our initial judgment is correct (even where we have certainty regarding relevant empirical facts) are precisely those for which an ethical theory is valuable.⁷ After all, unless we are willing to let our theory guide our judgments in at least some cases, working out an ethical theory is just an academic exercise. Another reason to exclude an initial belief from the set of considered judgments is that we have reason to think that the belief results from some distortion in our thinking. For example, someone who is having an affair may have a vested interest in concluding that adultery is not wrongful and this might bias their judgments.⁸ Other potential distorting factors include that the judgment is made in a hurry, that it is made while angry, that the person making the judgment has a close relationship with one of the parties to a conflict, and so forth. These are all reasons to exclude individual initial beliefs that reflect our intuitive judgments. In Section 2.4, we consider more wholesale objections to the use of intuitions in moral theorizing.

After weeding out the initial beliefs whose credibility we have reason to doubt, we are left with a set of considered judgments that consists of judgments about individual cases and about principles of varying levels of generality. This set is the data with which we try to construct a theory about the topic that interests us, whether it is a theory of the ethics of paternalism or a complete moral theory. Typically, the set of considered judgments will not be sufficient to specify our theory completely. This is for two reasons. First, there will usually be some inconsistency among the members of the set and so some adjustment is needed. One basic criterion for coherence among a set of beliefs, and one of the most basic virtues of a theory, is that it be internally consistent – that is, free of logical contradictions. Second, our choice of ethical theory will still be *underdetermined*

⁶ Since we will have different degrees of confidence in our initial beliefs, strictly speaking, we will require more evidence to reject some than others. The dichotomy that we describe here between considered judgments and initial judgments that we reject is a simplification that is useful for explanatory purposes.

⁷ As John Rawls explains: “There are questions which we feel sure must be answered in a certain way. For example, we are confident that religious intolerance and racial discrimination are unjust. We think that we have examined these things with care and have reached what we believe is an impartial judgment not likely to be distorted by an excessive attention to our own interests. These convictions are provisional fixed points which we presume any conception of justice must fit. But we have much less assurance as to what is the correct distribution of wealth and authority. Here we may be looking for a way to remove our doubts” (*A Theory of Justice*, revised ed. [Cambridge, MA: Belknap, 1971/1999], 17–18).

⁸ Konrad Bocian and Bogdan Wojciszke, “Self-Interest Bias in Moral Judgments of Others’ Actions,” *Personality and Social Psychology Bulletin* 40 (2014): 898–909.

by our set of considered judgments even when they are consistent – meaning that multiple theories will be consistent with the same set.

A great deal of debate in bioethics involves looking for and exposing apparent inconsistencies. For example, suppose we are interested in the conditions under which consent is valid and have agreed that voluntariness is one such condition. We are now developing a theory of what makes an act (such as giving consent) voluntary or involuntary. A *prima facie* plausible principle might be “Someone acts involuntarily if they are caused to act by someone or something external to them.” Now someone suggests this counterexample: if someone offers me a reasonable hourly rate to tutor them and I agree to do so, then I have been caused to act by something external to me (the prospect of money and satisfying work), but this is surely a voluntary act. After all, if it were not voluntary, then my consent to receive the money would be invalid, and that seems highly implausible. The structure of this simplified dialectic is as follows. We have a principle that was initially part of our set of considered judgments. A case was proposed and a moral verdict rendered about that case (that the action was morally unproblematic and thus voluntary). The case judgment appeared inconsistent with the principle. In such a case, resolving the inconsistency requires rejecting the principle, rejecting our intuitive verdict about the case, or some argument to show that we were mistaken about their inconsistency.

In the process of reflective equilibrium, decisions about how to resolve inconsistencies are very important. In the case just described, we expect that most people would be inclined to reject the principle: our intuitive verdict on the counterexample is one in which we have confidence; similar counterexamples seem likely to arise for many familiar cases in which someone is caused to act; and it seems likely that the principle was oversimplified. The natural course to take is to try to articulate another principle that is intuitively plausible without being subject to such counterexamples. But it will not always be obvious which of our considered judgments should be rejected. For principles in which they have more confidence, people may be inclined to preserve the principle and reject the judgment that called that principle into question. This kind of “biting the bullet” is common among philosophers and bioethicists who are seeking to challenge received wisdom and make what they consider to be moral progress. For example, in Chapter 7, our examination of moral status leads us to reject common intuitions about ways in which it is permissible to treat nonhuman animals and preserve the principle that the well-being of all sentient creatures has substantial moral importance.

The underdetermination of theory by data is a long-standing challenge for the development of scientific theories that also applies to theory choice in ethics.⁹ Here is a simple version of the problem. Suppose you are collecting empirical data in order to develop a scientific theory. For any finite data set – and all actual data sets are finite – there are infinite functions that would yield those data. This means that there are infinitely many universal generalizations that are consistent with the data. Which we should pick as our scientific theory for the phenomenon being studied is simply not determined by the data alone. Identical points apply to the construction of an ethical theory through the back and forth of the method of reflective equilibrium: the set of considered judgments will not determine which moral theory we should adopt.

Other Theoretical Virtues

The issue of how to resolve inconsistency and the underdetermination of theory by data both imply that the decision about what ethical theory we should adopt must be made on the basis of more than simply asking which theory is consistent with our considered judgments. Logical consistency is only one theoretical virtue. When we compare competing theories, we have to consider others.

One such virtue is the *prima facie* plausibility of the theory itself. Are the principles that make up a theory themselves ones in which we have a great deal of confidence or are they dubious? For example, many utilitarians find the theory compelling because its basic principle – that individuals should act so as to bring about the greatest overall improvement in well-being – seems so clearly correct to them. By contrast, for many people a moral theory based on the principles articulated by the biblical ten commandments would be implausible in part because it includes principles (e.g., “Thou shalt not steal”) whose exceptionless character they find dubious.

A second important virtue is the explanatory power of a theory. An ethical theory has greater explanatory power when it renders verdicts in more types of case than a competitor. We can assess this in two ways. First, one theory may be better able to give a verdict because it is more precise than another. So, for example, a theory that relies on intuitively weighing

⁹ Kyle Stanford, “Underdetermination of Scientific Theory,” in Edward Zalta (ed.), *Stanford Encyclopedia of Philosophy* (Winter 2017 edition; <https://plato.stanford.edu/entries/scientific-underdetermination/>).

competing principles will have less explanatory power than one that explicitly says how competing considerations should be balanced. Second, one theory may have broader scope than another, in the sense that it applies to more areas of our moral lives. For example, a theory of consent that is applicable in the domains of clinical research, sexual relations, and contract law has greater explanatory power than one that is tailored solely to clinical research.

Theories with greater explanatory power are more informative since they are able to provide moral verdicts for a wider range of cases. This also means that they are more open to counterexamples. If one theory is more precise than another, then it will be easier to see what it implies. It will be a “clear target,” making it easier to identify an implication that is inconsistent with some considered judgment. Likewise, if one theory has broader scope than another, then the first theory is more liable to being inconsistent with considered judgments in the form of principles or case judgments regarding one of the varied domains to which it applies. When we are comparing two theories we therefore need to be careful that we are not rejecting one that is more precise or has broader scope simply because it is easier to identify potential counterexamples to such a theory.

An important test for a theory occurs when it is extended to unfamiliar cases. It is evidence in favor of a theory if it renders verdicts about those cases that are also intuitively plausible – that is, entails moral verdicts about unfamiliar cases that are independently excellent candidates for considered judgments. It is a problem for a theory when its implications for novel cases conflict with considered judgments. In the face of such inconsistency one can adjust one’s theory to take account of the apparent counterexample. Such adjustments can then make the theory more or less informative. It will be more informative if we can now apply it to a further range of cases to test how it fits them. It will be less informative if the adjustment simply deals with the problematic cases, but no more. Adjustments like the latter are *ad hoc*. For example, return to our principle concerning voluntariness. We might adjust it to say: “Someone acts involuntarily if they are caused to act by someone or something external to them that they do not endorse.” Or we might adjust it to say: “Someone acts involuntarily if they are caused to act by someone or something external to them unless they want a tutoring job.” The latter is *ad hoc* – it generates almost no new predictions to test against. The former is much more informative – we can now examine various cases of endorsement to see how well the principle answers questions about voluntariness.

A final, related theoretical virtue is simplicity. It is generally thought that if two theories have the same explanatory power but one is derived

from fewer or more concisely stated principles, then the simpler one is better. Both utilitarianism and Kantian ethics, for example, might be regarded as simple in this sense, since they (purportedly) derive all their moral verdicts from just one principle applied to the empirical facts of a case. It is widely accepted that simpler *scientific* theories are preferable. For example, in addition to its greater explanatory power, one advantage that Newton's laws of motion and gravitational attraction had over prior physical theories was that their explanation of the movements of celestial bodies was simpler. What justifies this preference for simplicity and whether it applies equally in ethics has received little theoretical attention.¹⁰

The goal of the method of reflective equilibrium is to develop a moral theory that preserves as many of our considered judgments as possible, while remaining logically consistent, independently plausible, explanatorily powerful, and simple.¹¹ Naturally, there are trade-offs to be made. For example, we may find ourselves caught between a complex theory with many different principles that captures most of our considered judgments about cases and a theory that is much simpler but which requires us to amend more considered judgments. How to trade off the different theoretical virtues is itself a matter of debate, in ethics as in science.¹²

Reflective Equilibrium and Practical Ethics

The discussion so far may seem rather abstract and distant from ordinary ethical problem-solving. After all, when we are trying to decide what to do – in the clinic or outside it – it does not seem as if we are gathering a set of considered judgments and then constructing a theory from it. However, we think that the method of reflective equilibrium is implicitly used in everyday ethical debates and problem-solving. Consequently, understanding the method will help us adjudicate these debates and problems.

¹⁰ For a discussion of the challenges involved in justifying and applying simplicity criteria in the context of scientific theory choice, see Alan Baker, "Simplicity," in Zalta, *Stanford Encyclopedia of Philosophy* (Winter 2016 edition; <https://plato.stanford.edu/entries/simplicity/>).

¹¹ An ethical theory should, of course, also be consistent with what we have most reason to believe about relevant factual matters, including the best natural and social scientific theories. For example, an ethical theory whose scope was limited to human beings because it assumed that no nonhuman animals were conscious or sentient would be deficient due to its implausible factual assumption about animals. And a political theory that assumes that its preferred form of government is most conducive to human happiness is no stronger than this pivotal empirical assumption.

¹² Thomas Kuhn, "Objectivity, Value Judgment, and Theory Choice," in Alexander Bird and James Ladyman (eds.), *Arguing about Science* (New York: Routledge, 2013), 74–86. See also Ben Sachs, *Explaining Right and Wrong* (New York: Routledge, 2018).

First, even when only a narrow topic area is at issue we can often understand a debate in terms of reflective equilibrium. For example, consider what might get brought up in a discussion about the ethics of medical assistance-in-dying (MAiD). The discussants will want to show that their ultimate views are consistent with more general moral principles that they hold. Someone might invoke the importance of the right of competent adults to decide what happens to their bodies, or a physician might note the apparent incongruity between causing death and the role of healer. The resulting back and forth might involve amending their views on MAiD; it might also involve changing how they interpret those more general principles.¹³ Someone's views may also be challenged by showing that they appear to be inconsistent with a considered judgment about a case. For example, someone who thinks that it is permissible for clinicians to let someone die but not actively to kill might be confronted with a case in which that distinction does not seem to affect her moral verdict. This is the intended effect of James Rachels's fictional description of two evil uncles: both intend to murder their nephews by drowning them in the bath, but only one carries out his scheme, since the other has the "good fortune" to witness his nephew slip and fall and so only has to watch while he drowns.¹⁴ When one of the people discussing MAiD reflects on these apparent inconsistencies and decides how to respond, she will then have to make use of the considerations we described above. For example, she may be pushed to distinguish those judgments in which she is truly confident (such as that it would be unethical to kill a competent adult against his wishes) from those in which she is uncertain (such as whether it could be permissible for a physician to give a lethal dose to a patient who requests it). For these latter cases she may be seeking guidance from a theory.

Second, when we are debating about ethics – or when we are simply trying to give someone advice – we have to use something like the method of reflective equilibrium if we are to proceed in an effective, mutually respectful way. I can only persuade you of my view about some topic if I start from what you already believe in and show that given your beliefs it is reasonable to draw the same conclusions that I have. For example, suppose that one person is trying to persuade another that he should not eat pork. She might try to show him that eating pork is inconsistent with being a good Muslim. But if he is not religious, this will not be persuasive

¹³ See Chapter 4.

¹⁴ James Rachels, "Active and Passive Euthanasia," *New England Journal of Medicine* 292 (1975): 78–80.

because he lacks the requisite beliefs. Alternatively, she might ask him whether it is bad to cause humans to suffer. Perhaps he agrees. She might go on to quiz him about whether he can think of a reason why the suffering of humans matters but the suffering of other intelligent mammals does not. Perhaps he cannot. Finally, she may ask whether it is justifiable to cause another to suffer in order to gain a small amount of pleasure and he may agree that it is not. Then, if she presents him with data on how the pigs from which his pork comes are treated, he may be compelled to agree that he should not eat pork. Of course, this dialectic is simplified, but we hope it is recognizable. In starting from where the other person is already, it is possible to persuade them of an ethical view that they did not originally hold. The process of doing so essentially involves showing them that making their set of ethical beliefs optimally coherent requires accepting that ethical view. It is the method of reflective equilibrium.¹⁵

2.3 Alternative Methodologies

Philosophers and bioethicists have articulated a variety of methods for normative theorizing in bioethics. These methods, such as principlism and casuistry, were articulated within academic bioethics as rivals to one another.¹⁶ For principlists, such as Beauchamp and Childress, the application of mid-level principles to cases is intended to supply guidance as to what to do in those cases. Casuists, on the other hand, contend that the attempt to answer bioethical questions by applying agreed-upon principles to cases fails to take account of the rich contextual details that matter for actual decisions. Instead, bioethicists should proceed by careful description of the case under discussion and analogical reasoning from paradigm cases about which we have confident ethical judgments.¹⁷

With a couple of exceptions we think that these are all variants of the method of reflective equilibrium that differ in terms of the relative emphasis that they put on different types of considered judgments. For example, it is not true that casuists refuse to theorize at all. They have to make some

¹⁵ The same process occurs in written work on applied ethics. The writer attempts to show the reader that she should accept the conclusions for which he is arguing *on the assumption* that the considered judgments he cites are shared by her.

¹⁶ John Arras, "Theory and Bioethics," in Zalta, *Stanford Encyclopedia of Philosophy* (Winter 2016 edition; <https://plato.stanford.edu/entries/theory-bioethics/>).

¹⁷ For a history and defense of casuistry, see Albert Jonsen and Stephen Toulmin, *The Abuse of Casuistry* (Berkeley: University of California Press, 1988).

generalizations in order to draw analogies between similar cases and to decide which features of those cases are in fact relevantly similar.¹⁸ Rather than being simply opposed to universal principles, modern casuists may be understood as putting greater emphasis on the evidentiary weight of case judgments and accepting complexity in their universal principles as the price of ethical accuracy. Scholars who are more sympathetic to principlism, on the other hand, may be characterized as putting more weight on the importance of bringing cases together under universal moral principles. Again, such scholars do not typically deny that their theory should be sensitive to contextual details or to strongly held judgments about cases. Thus, these different methods simply vary in the importance that they attach to the different theoretical virtues described in the previous section.¹⁹

Bioethicists at either extreme of the methodological spectrum could deny that they are engaged in the method of reflective equilibrium. At one extreme, some particularists deny that moral principles are a source of justification. At the other, some foundational moral theories deny that considered judgments about cases and mid-level principles have any justificatory weight. We now argue against these possibilities in turn.

Some proponents of particularism claim to reject the use of theory altogether. For example, Jonathan Dancy denies that moral principles have any justificatory weight at all: “Moral Particularism . . . is the claim that there are no defensible moral principles, that moral thought does not consist in the application of moral principles to cases, and that the morally perfect person should not be conceived as the person of principle.”²⁰ Dancy argues that the moral relevance of any feature varies across cases such that, depending on the situation, the same feature may be morally good, bad, or simply neutral. Pain, for example, is bad in some situations – such as for a patient seeking treatment for his arthritis – but can be good in

¹⁸ Albert Jonsen, “Casuistry: An Alternative or Complement to Principles?,” *Kennedy Institute of Ethics Journal* 5 (1995): 237–251.

¹⁹ For more extended arguments for the claim that proponents of these different methods are all engaged in versions of reflective equilibrium see, e.g., John Arras, “The Way We Reason Now: Reflective Equilibrium in Bioethics,” in Bonnie Steinbock (ed.), *Oxford Handbook of Bioethics* (New York: Oxford University Press, 2007); Mark Kuczewski, “Casuistry and Principlism: The Convergence of Method in Biomedical Ethics,” *Theoretical Medicine and Bioethics* 19 (1998): 509–524; and Daniels, “Wide Reflective Equilibrium in Practice.”

²⁰ Jonathan Dancy, “Moral Particularism,” in Zalta, *Stanford Encyclopedia of Philosophy* (Fall 2017 Edition; <https://plato.stanford.edu/entries/moral-particularism/>). For an extended defense, see Jonathan Dancy, *Ethics without Principles* (Oxford: Oxford University Press, 2006).

others – such as when felt by athletes striving to push themselves as hard as they can. Likewise, pleasure is usually good, but can be bad – as when a sadist takes pleasure in another’s pain. Principles, such as “Pain is bad” or “Pleasure is good,” seem inevitably to be vulnerable to counterexample. Particularists like Dancy think that we can abandon them and simply explain our moral judgments by reference to the reasons that are relevant in each particular case, without the expectation that those reasons will operate in the same way in other cases.

Dancy’s view has been subject to extensive philosophical critique elsewhere.²¹ Instead of recapitulating that debate here, we note two key points. First, for extreme particularists like Dancy, we should demand a high burden of proof. If his view were correct, it not only would undermine the methodological points we made above about selecting a theory but also would require us to revise our everyday practices of discussing and teaching morality, since they often seem to involve searching for, demanding, and articulating moral principles.²² Second, insofar as moral particularism is supported by the apparent counterexamples that can be raised to proposed universal principles, so can the contrary view be defended by arguing in favor of specific universal principles. If a purported principle explains our considered judgments and gives us plausible verdicts for cases about which we are uncertain, that is a reason to preserve the principle. The arguments about principles that constitute the majority of this book stand as an attempt to demonstrate this point. We leave it to the reader to decide whether our theorizing is fruitful.

At the other extreme from the particularist position are views that seek to derive their answers to questions of applied ethics from foundational ethical theories, where the evidence for the truth of those theories is independent of how well they fit with more granular considered judgments about principles or cases. For example, Immanuel Kant sought to derive all of morality from the Categorical Imperative, which itself is a principle of rationality for beings like us (that is, embodied and able to act according to reasons).²³ Likewise, some utilitarians reject intuitive judgments as a

²¹ See discussions in Mark Lance, Matjaž Potrč, and Vojko Strahovnik (eds.), *Challenging Moral Particularism* (Oxford: Routledge, 2008).

²² Cf. Margaret Little, who writes: “If we reflect on our shared moral life, it certainly looks as though an important part of how we justify, convince, teach, and clarify is by pointing to explanatory generalizations whose truth we seem to endorse” (“On Knowing the ‘Why’: Particularism and Moral Theory,” *Hastings Center Report* 31 [4] [2001]: 32–40, at 36).

²³ *Grounding for the Metaphysics of Morals*, 3rd ed., trans. James Ellington (Indianapolis, IN: Hackett, 1993; first published 1785).

source of evidence about morality.²⁴ For such foundationalists, it might seem as though reflective equilibrium is irrelevant: the foundational moral theory justifies verdicts about cases, but verdicts about cases do not provide evidence for or against the foundational moral theory.

Like many others, we have yet to be convinced by a theory that attempts to derive all of morality from a single, allegedly self-evident principle. More importantly, for our point about methodology, one of the main reasons we find them unconvincing is that they fail to give plausible verdicts about cases. For example, one criticism of utilitarianism is that it implies that only the amount of benefits and harms matters, not their distribution. On its face, it therefore suggests that it could be permissible to punish an innocent person to calm an angry mob, or to ignore the needs of people who are severely disabled because it would be so expensive to benefit them. These implications are highly counterintuitive. This counts against any version of utilitarianism that has such implications.

2.4 Reflective Equilibrium: Clarifications and Criticisms

Why Start from Here?

One objection to the use of the method of reflective equilibrium is to ask why we should give any credence at all to our initial set of moral beliefs. What makes us think that starting with the moral judgments we are already disposed to make will lead us to end up with an accurate moral theory?²⁵ Given that the method of reflective equilibrium seeks to find a theory that preserves our considered beliefs, it seems plausible that one's starting point will bias where one ends up. For example, if you and I start with very different initial moral beliefs, then we are also likely to end up with different moral theories; that is, our reflective equilibria will be different. But why should I think that my starting point is preferable to yours? If I have no reason to think one starting point preferable, then I have no reason to think that one reflective equilibrium is preferable to another either. Skepticism seems to loom.

One possible response would be to claim that the method of reflective equilibrium, properly applied, will in fact lead to convergence between

²⁴ See, e.g., R. M. Hare, "Rawls' Theory of Justice" (in two parts), *Philosophical Quarterly* 23 (1973): 144–155, 241–252; Peter Singer, "Sidgwick and Reflective Equilibrium," *Monist* 58 (1974): 490–517; and R. B. Brandt, *A Theory of the Good and the Right* (Oxford: Clarendon, 1979), chap. 1.

²⁵ For blunt criticism along these lines, see Hare, "Rawls' Theory of Justice" (Part 1).

people who start with different moral views. Although we think this will be true in some cases – after all, a central point of moral deliberation is resolving disagreement – it seems unduly optimistic to think that this will always be the case. Further, for our skeptic, such convergence on its own might not be reassuring. The problem is not the possibility that we fail to reach agreement; rather, the problem is that our end point seems determined by our starting point and we have no reason to think that the starting point is correct. The possibility of two people coming to different equilibria because they have different starting points simply illustrates this worry. Thus, for the skeptic, convergence would be reassuring only if there were a plausible explanation of the convergence, for example, that the method of reflective equilibrium tracks reasons for belief and so brings us closer to moral knowledge.

At this point it is helpful to distinguish different objectives that we might seek with our methodology. If we want a method that will get us to the moral truth, then we need first to answer the deep questions in metaethics regarding whether moral claims can be true or false, what moral properties are, and how we come to know them. Depending on our answers to these questions, the method of reflective equilibrium may or may not prove to be the best way to access the moral truth. As we explain in Section 2.5, though we think there are strong grounds to reject moral skepticism, we do not have answers to these difficult and highly contested metaethical questions. We therefore regard the function of our methodology as more modest. The method of reflective equilibrium might not tell us how to get to the moral truth. Instead, it guides us to what we should say about novel or difficult moral questions, *given what we already believe*. Thus, it should not be seen as a response to moral skepticism, since it starts from the assumption that in a wide range of situations we already know what we should do. Similarly, with regard to interpersonal reflective equilibrium, we should be modest about what can be shown. It might be that people who start from very different views will not converge in their views on some subjects, even if they are the most patient and well-meaning of interlocutors. We can only attempt to convince those people who already share certain beliefs with us, that, *given those shared beliefs*, they have good reason to draw the same conclusions as we have for some novel or difficult question.

Even if our critic allows that there is no way of engaging in moral theorizing that is entirely independent of one's existing moral beliefs, it might be objected that the method of reflective equilibrium is still liable to give conservative results. After all, it involves trying to find the theory that

preserves as many of our considered judgments as possible. Since we start with the ethical beliefs that we (and, we hope, our readers) already have, we therefore stack the deck in favor of a moral theory that is similar to what we already believe.

But even brief reflection on the dominant moral views in Western societies over the last couple of centuries suggests that there have been dramatic changes in what many people believe rather than a conservative preservation of moral outlook. Moreover, it is hard, from our modern perspective, to avoid thinking that many of these changes constitute progress. For example, the prevailing views about people of different races or about women have not only changed, but surely changed for the better. A little humility suggests that there are likely to be equally dramatic changes in the future (perhaps concerning our treatment of nonhuman animals, for example).²⁶

Further, we would argue that the moral progress that has been made has occurred because of – not in spite of – the moral beliefs that people already hold. It is by realizing that certain of our beliefs are in tension with each other, that some are propped up by false empirical claims, or that some are clearly self-serving that the societal consensus has been pushed toward radical change. For example, a view that denies that women have the same moral status as men is one that is flatly inconsistent with most people's views about what underlies moral consideration (whether it be rationality, the ability to suffer, or species membership). The push for consistency between moral principles and moral judgments has made that view untenable.²⁷ Thus, although it is true that we start from where we already are, that fact does not prevent progress.

Empirical Concerns about the Reliability of Moral Intuitions

Recent empirical findings about how people's moral intuitions are elicited have also led some to skepticism about the role of intuitions in justifying moral principles. For example, Eric Schwitzgebel and Fiery Cushman describe a series of experiments in which they present participants with pairs of moral scenarios relating to the doctrine of double effect, the action-

²⁶ Cf. Peter Singer, *The Expanding Circle* (Princeton, NJ: Princeton University Press, 2011).

²⁷ For classic treatments, see John Stuart Mill, *The Subjection of Women*, vol. 1 (London: Longmans, Green, Reader & Dyer, 1869); and Simone de Beauvoir, *The Second Sex*, trans. Constance Borde and Sheila Malovany-Chevallier (New York: Random House, 2009; first published 1949).

omission distinction, and moral luck.²⁸ They show that the order in which the scenarios are presented has significant effects on moral judgments about the scenarios. Since order is presumably irrelevant to the right answer in these scenarios, the experiments cast doubt on whether intuitive judgments are a source of evidence about right and wrong. Joshua Greene and colleagues have conducted multiple experiments looking at variants of trolley problems.²⁹ They argue that people's intuitive responses are highly sensitive to the use of personal force. Since we do not think that the mere fact of using personal force rather than something else (e.g., pushing someone off a bridge rather than using a remote switch to drop him through a trapdoor) is morally relevant, they argue that we should not trust these intuitions.

For some philosophers, such findings throw the whole method of reflective equilibrium into doubt. For example, Peter Singer argues:

At the more general level of method in ethics, this same understanding of how we make moral judgments casts serious doubt on the method of reflective equilibrium. There is little point in constructing a moral theory designed to match considered moral judgments that themselves stem from our evolved responses to the situations in which we and our ancestors lived during the period of our evolution as social mammals, primates, and finally, human beings.³⁰

We agree that empirical findings about the origins of our moral beliefs and the causes of our moral judgments should be taken seriously. However, we think that the method of reflective equilibrium, as we have described it, is able to incorporate their use. For example, if our moral intuitions about some family of cases are highly sensitive to morally irrelevant features of those cases, we agree that this gives us reason to question the evidentiary value of those intuitions (so they should not enter the set of considered judgments). Thus, scientific evidence can play a helpful debunking role.

²⁸ Eric Schwitzgebel and Fiery Cushman, "Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-philosophers," *Mind & Language* 27 (2012): 135–153.

²⁹ See, especially, Joshua Greene, "Beyond Point-and-Shoot Morality: Why Cognitive (Neuro) Science Matters for Ethics," *Ethics* 124 (2014): 695–726. A "trolley problem" typically involves a runaway trolley and a choice between five people dying and one person dying. Variations in the nature of the choice include throwing a switch to move the trolley from one track to another, pushing someone off a bridge to block the trolley, and so forth.

³⁰ Peter Singer, "Ethics and Intuitions," *Journal of Ethics* 9 (2005): 331–352, at 348. But see, e.g., p. 347 on the possibility of a wide reflective equilibrium that could "countenance the rejection of all our ordinary moral beliefs." For a more extended response to Singer with which we are sympathetic, see Joakim Sandberg and Niklas Juth, "Ethics and Intuitions: A Reply to Singer," *Journal of Ethics* 15 (2011): 209–226.

However, it can only play this role along with considered normative judgments. The judgment that some feature of a case (e.g., the order in which cases are presented) is morally irrelevant is also a considered judgment that we employ in the debunking argument. Even the most hardcore skeptics about the evidentiary value of intuitions acknowledge this general point.³¹

Furthermore, we believe that the available evidence does not impugn the majority of careful work in applied ethics that makes use of judgments about cases. A great deal of this work does not rely on brute intuitions – like a gut response that I should not push someone off a bridge – but uses cases to draw out the structure of moral principles that we already have. For example, analyses of coercion, consent, or the nature of prudential value appeal to complex concepts with which many people are already facile. Take an example from theoretical work on consent. A. John Simmons describes a case in which the chair of a board asks attendees at a meeting if they have any objections to the policy he proposes.³² Their silence, Simmons points out, constitutes consent to the policy provided that it meets the same standards for voluntariness and the like that affirmative consent would require. But the reader who is persuaded by Simmons that “tacit consent” is morally transformative in the same way as express consent does not have a gut response to the case and conclude that Simmons has given an explanation. Rather, Simmons uses the case to illustrate a view the reader already endorses.

In summary, we welcome the empirical evidence, consider it relevant, and believe it should be used during the process of seeking reflective equilibrium along with the other relevant considerations we have described.

2.5 *Metaethics*

Work in metaethics involves the attempt to understand the ultimate foundations of ethics. Are there matters of fact regarding ethical

³¹ Some of these writers think that there are different types of intuition in play here. See Singer, “Ethics and Intuitions,” 350–351; and Greene, “Beyond Point-and-Shoot Morality,” 724, citing Sidgwick on “philosophical” intuitions. We believe they underestimate the number of such intuitions, which should include intuitions not just about consequences but about the nature of those consequences (e.g., suffering, death), about their distribution, and, as noted, about moral relevance.

³² A. John Simmons, “Tacit Consent and Political Obligation,” *Philosophy & Public Affairs* 5 (1976): 274–291, at 278–279.

judgments? Can such judgments be true or false, objectively correct or incorrect? If ethics admits of truth or objectivity, in what is it grounded: religious truths, some other type of metaphysical truths, facts about the natural world? How can we know the relevant facts? And so on.

These are enormously complicated matters that have been debated at least since antiquity.³³ They are not matters about which this book has much to say. Nevertheless, some of our readers might wonder how we can have a theory of bioethics without addressing them. In the following paragraphs we sketch answers to some of the questions such readers might have.

What are you assuming about the nature and foundation of ethics in using the method of reflective equilibrium? Our assumptions are relatively modest. We are not committed to any specific view of the foundation of ethics. In fact, we do not seek a rationally indubitable foundation for ethics and doubt that such a foundation exists. Further, we assume nothing about the truth or falsity of particular religions or religion in general. We consider it inappropriate to appeal to the supposed authority of some individual, a particular group, or a religious text as the basis of ethical thinking. We do assume that people's beliefs about ethical matters, especially upon reflection and when informed about relevant facts, provide the appropriate starting point for ethical inquiry. In the absence of an indubitable foundation or infallible source of authority for ethics, we think, there is no more credible starting point than what people believe about ethics.

Are you assuming that ethical beliefs can be true or false, that there are facts of the matter regarding ethical issues? The answer may depend on how broadly, or narrowly, one defines "truth" and "facts" – and we do not wish to enter this semantic territory. What we can say is that we assume that ethics is objective in at least the sense that there are better and worse answers to ethical questions, that some ethical judgments are more defensible and worthy of acceptance than others. Without such an assumption there would be little or no point in investigating and debating ethical issues. Why do so if no result is better than any other? We therefore reject *ethical skepticism*, which holds that no ethical judgments are justified and therefore better than any others, period. We also reject *ethical relativism*, which (as we understand it) holds that ethical judgments can be justified only relative to the ethical beliefs of a particular culture or group.

³³ For a useful collection of historical and contemporary readings, see Steven Cahn and Andrew Forcehimes (eds.), *Foundations of Moral Philosophy* (New York: Oxford University Press, 2016).

Why do you reject ethical skepticism? Our confidence has several grounds, which we can present only briefly here. First, we find certain ethical judgments – and the belief that they are binding on all human moral agents – more plausible than any arguments we have encountered in support of ethical skepticism. For example, we find the judgment “Raping children is wrong,” where this judgment is understood to apply to all human beings, far more plausible than the argument that, because there is no God, everything is morally permissible. Likewise for every other argument we have encountered in support of ethical skepticism. If we accepted the conclusion of the argument, then we would have to accept that all our ethical beliefs are mistaken, and this seems more counterintuitive than that the argument is unsound. Similar points apply to examples of apparent moral progress. Increasing respect for gay persons in Western countries in recent decades seems to represent an ethical advance over the comparative disrespect that preceded it. Unless ethics were objective in the sense that there are better and worse answers to ethical questions, there would be no standard against which we could measure the trend of increasing respect as an improvement rather than simply a change.

Second, it is very difficult to maintain skepticism about ethics without also becoming a skeptic about all reasons for action. Consider a very basic ethical claim: that an individual agent should take the interests of others into account when deciding what to do. The skeptic says that this claim is false. The fact that some act will help or hinder another person is in itself irrelevant to what the individual should do. How should an individual agent decide what to do? Perhaps, our skeptic might suggest, she should think only of her own interests and how to promote them. In that case, her interests provide her with reasons for action. But now we may ask why even her own interests provide her with reasons to act. Certainly, we humans are less prone to doubt that we have good reason to promote our own interests than other people’s, but that does not justify the claim that we should care about them. If we should be ethical skeptics, then perhaps we should be prudential skeptics too.

There are three possible ways to respond to this argument. The first is to embrace wholesale skepticism about reasons and say that no one has any reason to do anything. This is logically consistent but seems impossible for any actual agent to adopt. Whenever one faces a novel situation and stops to think about what to do, the decision process involves thinking about the reasons to do one thing rather than another. The second response is to show that there is a difference between prudential and ethical reasons such that we should accept the former but not the latter. This would require

some convincing explanation of why our own interests give us reasons, but the interests of others do not. The third, which we prefer, is to accept that there are both prudential and ethical reasons. Your interests matter to you, mine to me, and ours to each other. The challenge for ethical theory is to work out how they matter.³⁴

Why do you reject ethical relativism? One reason we do so is the same as for rejecting ethical skepticism: our confidence in some of our ethical judgments. We are confident that committing genocide and raping children is wrong. We have yet to hear the argument for relativism that is convincing enough to shake our conviction that these actions would still be wrong even if a particular culture or group believed otherwise.

A second reason is that ethical relativism does not have a satisfactory way to justify ethical claims to those who disagree with them. Suppose that someone grows up within a culture but comes to disagree with a commonly held view within that culture about gender roles. She finds the views of a different culture with more liberal gender norms more plausible. According to the ethical relativist, the fact that her culture has a specific ethical view is justification for that view: she is wrong to defy these gender norms. But, she may ask, how is it that these norms correctly apply to me but not to women in another culture, just because I grew up in one and not the other? The relativist must say that ethical judgments are ultimately justified just because the majority of people in a culture believe them. This justification seems unsatisfactory: the dissenter is asking for reasons why she should conform to cultural norms, not just the assertion that they are cultural norms – that is, judgments held by the majority.

Finally, some of the most commonly presented grounds in favor of ethical relativism actually support an objective understanding of ethics. For example, one often hears that we should be ethical relativists because it would be disrespectful to condemn the ethical systems of other societies when they differ from our own society's ethical views. This reasoning implies that respect for other cultures is ethically valuable. Yet surely such respect is not valuable only because our own culture says it is. Disrespect seems morally problematic, no matter who the disrespectful agent or culture is. Moreover, those who advance this argument in favor of relativism usually acknowledge limits to appropriate deference to other cultures' views. It is not as if respecting the views of another culture means we

³⁴ For longer and more sophisticated arguments in a similar vein, see Thomas Nagel, *The Possibility of Altruism* (Oxford: Clarendon, 1970). See also Derek Parfit, *On What Matters*, vol. 1 (Oxford: Oxford University Press, 2011), Part 1.

should tolerate, for example, genocide or slavery in another society. The good point the ethical relativist has in mind is that we should not assume that our culture is correct on all ethical matters on which there are differences among cultures. But this point is consistent with believing that there are objectively better and worse answers to ethical questions. As later chapters will make clear, we do not defer to or accept all of our culture's views on ethical matters. For example, we argue that the dominant Anglo-American culture is wrong in not viewing animals as having substantial moral status and in often favoring property rights over the most important needs of the global poor.³⁵

Consistent with the method of reflective equilibrium, we should take existing ethical beliefs (of anyone from any culture) to have some initial authority but not as infallible. This approach is appropriately respectful of members of other cultures without falling into the implausibility and impracticality that characterize ethical relativism and ethical skepticism.

This concludes our discussion of methodology in ethics. The task of the next chapter is to sketch our ethical theory.

³⁵ For an approachable overview of reasons why one might be tempted by ethical relativism and why that temptation should be resisted, see Michael Garnett, "Is Morality Relative?" (unpublished manuscript available at <https://philpapers.org/archive/GARIMR.pdf>).