

INTRODUCTION

Jussi Suikkanen

Final author copy; To appear in *Methodology and Moral Philosophy*, Jussi Suikkanen and Antti Kauppinen (eds.), Routledge.

Abstract:

This chapter begins by explaining two widespread attitudes towards the methods of moral philosophy. The first common attitude is that the appropriate method for doing ethics was described by John Rawls when he formulated the reflective equilibrium method. Another common attitude is that moral philosophy has no method – anything goes in ethical theorising as long as the results are significant enough. The chapter then motivates the volume by arguing that these attitudes are not helpful. The reflective equilibrium method has its limits and yet not all ways of proceeding in ethics are equally good. For this reason, I argue that we need to be more aware of the argumentative strategies we employ in ethics. This requires being methodologically reflective and transparent and taking part in the debates about the merits and problems of different methodologies exactly in the way done in the chapters of this volume. The second half of the chapter then provides an outline of the other chapters. Here I focus on clarifying exactly how these chapters contribute to the new discussions about the methods of ethics.

Background

I often come across two different but equally perplexing attitudes towards methodological questions in moral philosophy. Firstly, many think that, as a discipline, moral philosophy has a unique method, comparable in significance to the ‘experimental method’ of the natural sciences. This well-known *reflective equilibrium* method was first explicitly outlined by John Rawls (1951, 1974 and 1974). Many ethicists appear to think that knowing the basics of how the reflective equilibrium method works is all you need to know about how moral philosophy should be done.

The reflective equilibrium method is easy to outline. We must begin from judgments about individual cases which must be held sincerely and which must also be stable in the kind of careful deliberation that is not distorted by strong emotions or self-interested bias. In the second stage, we attempt to formulate a set of general moral principles that could both fit and also justify the previous convictions. When we formulate these principles at this preliminary stage, it will be likely that there will not be a perfect match between our carefully considered judgments about the cases and the general principles.

In the third stage, we then try to get rid of the previous conflicts in two ways. In some conflict cases, it makes sense for us to modify our judgments about the cases on the basis of the general principles because those principles support our intuitions so well elsewhere. In other cases, in contrast, it makes more sense to attempt to find new, more sophisticated principles so that we do not have to give up our convictions about the cases given how deeply held they are.

Finally, in the fourth stage, we fine-tune our principles by taking into consideration the leading ethical theories on the topic and the best arguments made in their support. We are also to seek wider reflective equilibrium by broadening the set of moral and non-moral beliefs with which the relevant moral principles we have formulated are supposed to be compatible. Thus, at this final stage, we check that the latter principles are compatible also with what we know, for example, about what kind of social systems can be stable and what influences our moral judgments.

It is easy to see just why the reflective equilibrium is such a promising account of how we ought to do moral philosophy. It is an ideal that seems provide us with almost something like an algorithm – clear steps which anyone should be able to follow systemically in order to achieve moral knowledge. The description of the method also intuitively seems to capture something about how many moral philosophers themselves understand what they are doing. It is thus not surprising that majority of moral philosophers continue to endorse the method at least in some form.

Despite this, many ethicists have always been torn about the reflective equilibrium method for several reasons. Firstly, there have always been many forceful objections to the method. One of the most poignant criticisms has always been that, even if when we use the reflective equilibrium method we must begin from our carefully considered moral judgments, these judgments are still bound to be a reflection of cultural indoctrination, superstition and bias (Hare 1975; Brandt 1979, pp. 21–2). As a consequence, whatever moral principles end up being in a reflective equilibrium with our carefully considered judgments, these principles will be just as unlikely to reflect the moral reality as the intuitions that serve as the input, or so the objection goes.

Even if this objection seems to have considerable force, it has not been able to change that many people's mind about the reflective equilibrium method. This is because, if we are not allowed to rely on our carefully considered judgments in the evaluation of moral principles, then we would be required to evaluate those principles from a completely non-moral, non-evaluative point of view.¹ Yet, in this situation, it becomes less clear on what grounds we could be able to choose between alternative

¹ Some have suggested that, instead of carefully considered judgments about cases, we can rely on carefully considered judgments about more general principles (see, e.g., Singer (2005)). This responses assumes, however, that there is an important epistemic difference between the two kinds of judgments.

moral principles (Hooker 2000, p. 11). As Frank Jackson (1998, p. 135) put it, ‘we must start from somewhere in current folk morality, otherwise we start from somewhere *unintuitive*, and that can hardly be a good place to start.’

The reflective equilibrium method also suffers from other problems – two of which are relevant here. Firstly, many ethicists feel like the description of the reflective equilibrium method is not especially helpful when it comes to their everyday work. Somewhat boringly, the method merely guides us vaguely to seek coherence between our carefully considered moral convictions, moral principles and the relevant empirical beliefs. Yet, very few people have thought that incoherent views are better than coherent ones and more importantly the reflective equilibrium method does not tell us exactly how we should make our principles cohere with our carefully considered convictions. In any given conflict situation, is it better to give up your carefully considered conviction or reformulate the general principles? The method itself does not tell and therefore, when we make these choices, we must rely on our judgment. Yet, nothing in the description of the method helps us to avoid making mistaken judgments.

The second issue is that very few recent works in moral philosophy employ the reflective equilibrium method in any recognisable and explicit form. One reason for this is that many ethicists work on questions that simply cannot be answered by using it. If you work in normative ethics and are searching for general principles that capture what kind of actions are right and wrong across different contexts, then at least in principle you can rely on the method. Yet, if you happen to be working on any of the other equally interesting and important questions in normative ethics or if you

are trying to solve problems in either metaethics or applied ethics, then there just does not seem to be any straightforward way in which you could rely on the reflective equilibrium method. For example, in metaethics finding a reflective equilibrium between our moral principles and carefully considered convictions about cases will tell us little about the meaning of moral concepts, the nature of moral judgments or the essence of moral properties. Likewise, in applied ethics, it is equally unlikely that the previous type of coherence between convictions and general principles could shed much light on which morally salient considerations should be taken into account when we consider the most difficult moral questions concerning the climate change, war, new biomedical technology and so on.

Perhaps these limitations of the reflective equilibrium method have led to another extreme in the ethicists' attitudes towards the methodological questions. The second common methodological attitude is a reflection of Paul Feyerabend's (1975) 'anything goes' view in the debates concerning the appropriate methods of science. The core idea behind this attitude is the observation that, if we consider both historical and more recent key contributions to moral philosophy, it is difficult to extract a unique method such that its use would explain the fact that these very contributions constitute the most important advances. Rather, what we seem to find from the key works of moral philosophy is very different kinds of arguments and so, in a sense, the biggest advances have been made by using very different kinds of methods. As a consequence, it just is not plausible that any strict methodological rules would have governed the growth of knowledge in moral philosophy any more than in the case of natural sciences. Furthermore, those who have this liberal methodological attitude emphasise that, given that we do not yet know how the new advances will be made in

the future, we should not require all ethicists to follow any strict methodological rules. After all, doing so would only prevent us from making important discoveries the making of which will also require inventing new methods. We should thus let different ethicists pursue different lines of inquiry as they see fit and just see what works, or so the argument goes.

There are several reasons for why this ‘methodological anarchism’ is not wholly satisfactory either. The main problem is that the attitude is based on a presupposition according to which different views in moral philosophy can be evaluated independently of the arguments provided in their support. If the evaluation of different theories could in this way be distinguished from the arguments that are provided in their support, it would make sense not to be too concerned about the methods of ethics. Any method would do as long as it leads to the right results.

It is, however, much more appealing to think that how plausible different answers to different questions in moral philosophy are is always intertwined with the question of how strong arguments have been provided for those answers. That is, in moral philosophy, evaluating a philosopher’s views and the kind of arguments she has provided for them always go hand in hand. This is because there are no ways of evaluating different claims in moral philosophy that would be independent of the arguments made for and against those claims. One good reason for thinking that this is the case is that, usually when two ethicists have a substantive disagreement over some significant question, they equally disagree about the arguments used for defending those views. And, if this is the case, then it makes no sense to think that any way of doing moral philosophy is appropriate as long as it leads to the growth of

moral knowledge. The ways of proceeding that yield better arguments for the defended conclusions must be better than others.

The ‘anything goes’ attitude can also have harmful consequences for how moral philosophers proceed in practice. It seems to suggest that, as moral philosophers, we do not need to be self-reflective: we do not need to think about the ways in which we pursue our research. It seems to tell us that it is enough that we try to answer our research questions the best we can without paying much attention to the question of whether the methods of inquiry we are using are sound. After all, if we happen to stumble upon the right answers, the methods we used can always be declared appropriate in retrospect in the light of the answers we came up with. Yet, surely, this is not the best way to proceed even for individual researchers.

Rather, it is much more plausible to think that, as practicing moral philosophers, we should pay serious attention to how we intend to find the answers to the ethical and theoretical questions that interest us. Taking our methods seriously requires us, for example, (i) to consider the kinds of arguments that have been used before in ethics – what kind of structures they have and what kind of premises and presuppositions they begin from, (ii) to keep track of the advances in the empirical sciences that touch on the topics of our research, (iii) to follow and take part in the methodological debates within moral philosophy and (iv) to aim at making methodological innovations ourselves. If these recommendations sound overly demanding, one good reason for following them is that the most successful ethicists have always also at the same time been methodologically both innovative and self-aware.

Let me then draw three lessons from the problems of the methodological attitudes outlined above. Firstly, given that moral philosophers investigate very different kinds of questions, it is unlikely that there will be any single method of ethics that could be used to make progress in all debates of moral philosophy. It is unlikely that the best ways to approach the second-order metaphysical questions about the nature of moral properties are the same as the ways in which we should think about first-order questions such as what duties we, as individuals, have towards non-human animals. This suggests that, when it comes to the methods of ethics, we should be methodological pluralists rather than monists. Even if it is not the case that anything goes, there will be a number of different and equally appropriate methods which ethicists should be able to rely on with confidence depending on what kind of questions they are investigating.

The second lesson we should draw is that it is unlikely that even the best methods of ethics will be simple step-by-step algorithms, which even a computer could be programmed to follow. It is true that, in the methodological debates, it is often assumed that, if moral philosophy had a method, it would have to be something like a set of instructions that anyone could take off the shelf and apply successfully. Yet, as the example of the reflective equilibrium method suggests, it is unlikely that any method of moral philosophy could be specified at the level of specificity required for making it a simpler algorithm. Because of this, I believe that it is more useful to think of the appropriate methods of ethics as methods in a much looser sense.

Perhaps we should then think of different types of argumentative strategies – different ways of providing support for views – as the genuine methods of ethics. That is, we

could think that different ethicists who rely on different kind of implicit assumptions in their work, whose explicit arguments begin from different starting points (be they the results of empirical sciences, their personal ethical convictions, or results of abstract philosophical *a priori* reasoning), and who employ structurally different types of arguments all use different methods of ethics, loosely understood. It is unlikely that these actual methods could be captured in terms of simple step-by-step prescriptions, because using these methods in new contexts will always require judgment. Yet, despite this, we should still be able to observe enough similarities between different types of arguments in order to be able to evaluate meaningfully different general ways of proceeding.

Finally, insofar as we can identify different methods in the previous looser sense, we should resist the idea that all ways of proceeding are equally good. Instead, as I suggested above, we should be reflective about the ways in which we do moral philosophy. We should pay attention to different kinds of arguments, we should take part in the evaluation of different methods and we should be aware of the methods we use and willing to defend them when challenged. This is especially true in ethics because, as we have seen, it is impossible to distinguish the plausibility of a view from the strength of the arguments provided for and against it.

In this situation it is fortunate that, despite how common the two attitudes described above are, a lot of important work has been recently done on the methods of ethics. Yet, perhaps exactly due to the strength and popularity of those attitudes, many of the recent contributions to the methodological debates have failed to receive the attention they deserve. Of course, the merits of the reflective equilibrium method continue to be

debated and similarly the question of whether the empirical methods of psychology and social sciences could be used to make new progress in moral philosophy is receiving a lot of attention.² Yet, much of the important methodological work tends to remain more hidden.³ One indication of this is that currently there are no overview articles or textbooks that students could use for familiarising themselves with the different ways of proceeding in ethics. Similarly there are very few professional venues in terms of conferences or bespoke publications that would bring together ethicists who take part in the different methodological debates. For these reasons, at the moment, it is unfortunately quite difficult to get a sense of how the discipline of moral philosophy understands its own methods.

Because of this, I organised two conferences on the ‘New Methods of Ethics’ at the University of Birmingham: one in September 2016 and one in January 2017. With the help of funding provided by the Arts and Humanities Research Council and the College of Arts and Law, I was able to invite some of the key contributors to the methodological debates in ethics to these conferences. The amount of submissions I received through an open call for abstracts and the quality of these submissions similarly showed how much interest there really is for exploring the methods of ethics. The purpose of the two conferences was thus to bring together ethicists both from different areas of ethics and from different methodological debates so that they could discuss together the basic question of how we should do moral philosophy.

² For comprehensive overviews and lists of references, see Daniels (2016) and Alfano and Loeb (2017).

³ Many recent debates concerning the methodology of moral philosophy have investigated conceptual analysis, intuitions about cases, linguistic judgments (about the felicity or truth of various attributions), parsimony, anti-theory, the idea that an adequate account must be internal to a practice, and other equally interesting topics.

The aim of this volume is to disseminate the ideas explored at the two conferences to a wider audience. My hope is that the material published here will encourage ethicists to become methodologically more self-aware, critical and innovative. I thus hope that the interesting methodological debates covered here will also receive more attention in the discipline more broadly. Finally, I also hope that the chapters of this volume will prompt others to attempt to map the different ways of proceeding in ethics more clearly and in a way that could be presented in an accessible way to those who are new to moral philosophy. I firmly believe that achieving these aims will help the whole community of moral philosophers to make more progress.

The chapters

The chapters of this volume have been structured under four categories. The first three chapters take part in one of the most intensive methodological debates in moral philosophy, which is about what role, if any, the empirical methods of psychology and social sciences should play in ethical theorising. After this, the next two chapters will boldly outline brand new methods: a new first-personal method for doing metaethics and a new method for first-order normative ethics based on the idea of comparing the authoritativeness of different normative standards. The three chapters of the Part 3 will then critically evaluate some of the most influential methods used in moral philosophy recently. Finally, the last two chapters of Part 4 will consider the question of what role, if any, should our first-order normative intuitions play in the evaluation of different metaethical views. That is, these chapters attempt to answer the question of whether doing first-order normative ethics is one of the methods we need to rely on when answering metaethical questions.

Part 1: The Prospects of Empirically Informed Ethics

At the heart of the recent passionate debates concerning the so-called empirical (or experimental) ethics have been certain general attempts to debunk our moral beliefs. These sceptical arguments proceed in two stages. The debunkers first attempt to describe certain general causal mechanisms responsible for how we came to have our moral beliefs (such as evolution). At this point, the debunkers have also emphasised the fact that our moral intuitions seem to be vulnerable to framing effects (Sinnott-Armstrong 2008), produced by a ‘quick and dirty’ emotional systems (Greene 2013), or widely grounded in reactions of disgust (Kelly 2011). The debunkers then argue that these features of our moral beliefs undermine whatever justification we had for them because the processes responsible for these beliefs turn out to be unreliable.

In ‘How to Debunk Moral Beliefs’ (Chapter 2), Victor Kumar and Joshua May construct a dilemma for the previous kind of general attempts to undermine our moral beliefs. According to them, if the debunkers focus on a belief-formation mechanism such as evolution that could be responsible for most of our moral beliefs, they will struggle to show that this mechanism is unreliable given how many of our moral beliefs can also plausibly be assumed to be true. To avoid this problem, the debunkers can focus on more local belief-formation mechanisms based on disgust and other specific emotional reactions or on processes vulnerable to framing. The debunkers have a much better chance of showing that the moral beliefs acquired in these ways may be undermined. Yet, the trade-off here is that no general sceptical conclusions follow – it is always open for those who are not sceptics to argue that most of our moral beliefs have not been formed in these ways. Kumar and May thus attempt to

argue that there cannot be general sceptical, empirically orientated debunking arguments that would be effective enough to challenge all our moral beliefs.

Yet, Kumar and May also describe a new empirically informed method that could be employed successfully in more local ethical debates. Here they begin from the observation that the so-called ‘consistency reasoning’ is very often used in both everyday life and in moral philosophy. This type of reasoning relies on the simple idea that we should treat like cases alike. When our consistency is questioned, we have two choices: we can either change our view about one of the cases or we can explain what the morally relevant difference between the cases is. Kumar and May then propose that one useful role which empirical research can play is that it can reveal to us what features of different situations make us treat them differently – to what we are in fact reacting. After empirical investigation has revealed to us what makes us treat the cases we are considering differently, we, of course, need to consider whether those factors really make a genuine moral difference. Yet, given that it is always possible that we are in fact tracking intuitively irrelevant features, empirical investigation can play a significant role in powerful consistency arguments, or so Kumar and May argue.

One standard method on which most ethicists rely at some point is testing whether different suggested moral principles fit our moral intuitions about various fanciful problem cases. These cases are intentionally unrealistic as their purpose is to enable us to focus on just few isolated features of the situations. Focusing on these features in artificial thought-experiments enables us to test whether our moral principles carve the joints of the moral reality at the right places in a way that would be difficult to do

in messy real-life situations. One famous example of this type is the Trolley case in which a trolley is about to hit five people but you have an option of redirecting it to kill only one person.

Yet, recently the epistemic value of these cases for ethical theorising has been challenged on empirical grounds (Schwitzgebel and Cushman 2012 and 2015; Wiegman et al ?). It turns out that, for example, the order in which these cases are presented affects what intuitions even professional ethicists have about them and even the moral principles they are willing to accept. If the intuitions of ethicists can be manipulated this easily, there is a worry that they lack a competence to make the kind of reliable judgments about these cases that could offer us an insight to which moral principles are correct.

In Chapter 3 ‘Who’s Afraid of Trolleys?’, Antti Kauppinen attempts to defend the standard methodology of normative ethics by responding to these debunking empirical concerns. Kauppinen first argues that philosophers tend to employ their special competences only in theoretical contexts in which they evaluate different ethical views and for this reason the previous experimental settings might not prompt them to make their best carefully considered convictions. Kauppinen also offers an alternative explanation of why the ethicists’ judgments about the cases can be influenced by superficial things such as the order in which the cases are presented. This is because, in the relevant cases, there are many morally salient considerations that need to be taken into account. The order in which the cases are presented can then influence the way in which ethicists weigh these complex factors. Finally, Kauppinen also considers the ways in which we can attempt to screen off the intuitions that do

not illustrate our best moral competences. This can be done, for example, by being cautious when relying on uncertain intuitions in theorising and by understanding ethical inquiry as a *social* process in which many of our responses to the individual cases continue to be challenged by others in a way that can be corrective.

The third chapter taking part in the methodological debates concerning the role of empirical sciences is Tyler Millhouse, Alisabeth Ayars and Shaun Nichols's 'Learnability and Moral Nativism: Exploring Wilde Rules' (Chapter 4). According to moral nativism, evolution has provided us with an innate 'moral rule acquisition capacity' that enables us to learn moral rules relatively quickly.⁴ Yet, given that this capacity also imposes certain structures on our moral thought, the nativists also accept that the capacity must constrain the kind of moral rules we can learn.

If the nativist hypothesis were true, this would have significant consequences for many debates in moral philosophy. For example, if we accept metaethical realism and think that moral facts are independent of human judgments, nativism seems to entail that moral facts could in principle be such that we, as human beings, might not be able to learn the moral rules that best capture those facts. Likewise, if we are not metaethical realists but rather think that moral facts are in some way mind-dependent, then the nativist hypothesis offers us a new empirical way of investigating what the moral facts are. In principle, in this framework, we should be able to rule out ethical theories by empirically investigating whether the moral rules they entail would be learnable by ordinary human beings. Thus, whatever metaethical view we accept, the nativist hypothesis seems to have fascinating consequences.

⁴ For the most systematic presentation of nativism, see Mikhail (2011).

Millhouse, Ayars and Nichols, however, argue against the nativist hypothesis based on the empirical research they carried out. They begin from a type of moral rules that do not exist anywhere in the world. Nativists tend to think that these rules have some structural features such that the absence of these rules can be explained by the fact that they fail to satisfy the constraints on moral learning set by the innate moral rule acquisition capacity.

Millhouse, Ayars and Nichols focus on what they call ‘the Wilde rules’.⁵ These rules both permit agents to bring about an outcome intentionally and forbid agents to bring about that very same outcome unintentionally (that is, they prohibit merely allowing the outcome to come about). Such rules do not exist in real life: the moral rules we have either permit (or forbid) bringing a certain outcome both intentionally and unintentionally or they forbid intentionally bringing about an outcome whilst permitting merely allowing the outcome to come about.

Chapter 4 then begins from the thought that, if the learning of any moral rules were made impossible by the innate structure of the moral rule learning capacity, the Wilde rules would surely need to be amongst these rules. Yet, the empirical evidence provided by Millhouse, Ayars and Nichols shows clearly that even Wilde rules can be learned naturally without any explicit instructions or explanations. This should make us sceptical about the nativist hypothesis. Given that we can learn Wilde rules, there is no reason to think that any moral rules could have structural features that would make them unlearnable because of our innate moral rule capacities.

⁵ They call these rules Wilde rules because of a quote often attributed to Oscar Wilde according to which ‘a gentleman never offends unintentionally’ where this rule is taken to imply that intentional offending is permitted.

In the studies carried out by Millhouse, Ayers and Nichols, the research subjects were asked to imagine a foreign culture with the gibberish rule ‘nib weigns’. The subjects were then shown examples of violations of the rule. Some participants were given only examples that were clearly actions, some only examples that were clear allowings, and some both. The subjects were then asked how likely it was that certain new cases of actions and allowings were also violations. The results were revealing. Firstly, it was shown that the participants who were shown only examples of violations that were allowings were naturally able to learn Wilde rules: that is, without any instruction or explanation the participants would no longer take actions to be violations of the relevant rule. However, the participants also showed a certain bias against such rules. The participants shown only allowing examples were more inclined to think that also actions were violations than those who were only shown action examples were inclined to think that also allowings were violations. Millhouse, Ayars and Nichols argue that this bias can be explained in the framework of Bayesian accounts of learning – the participants are biased against the Wilde rules because their prior expectations of what the rule in question is like has been shaped by the moral rules they have encountered.

This chapter makes a number of interesting methodological contributions. Firstly, it illustrates with an example what kind of an important role empirical research can play in moral philosophy. It shows that many interesting philosophical hypotheses, for example, about the way in which moral rules are learned can be empirically tested and so there are areas of moral philosophy in which empirical research definitely has a role to play. Furthermore, in the light of Millhouse, Ayars and Nichols’s work it does

not seem likely that empirical investigation to which moral rules we can learn could shed much light on which actions are right and wrong even if we assume some form of mind-dependence of moral facts in metaethics. After all, the empirical research carried out here suggests that we are flexible learners.

Part 2: New Methods

Two of the authors of this volume have been ambitious enough to outline genuinely new methods of ethics. In her ‘Metaethics from a First-Person Standpoint’ (Chapter 5), Catherine Wilson begins from the observation that, even if in first-order normative ethics the first-person perspective of the agent is often taken seriously (consider, for example, the demandingness objections to consequentialism), the nature of obligations in metaethics has so far been explored purely from the objective third-personal perspective. Thus, her aim is to explore whether new progress could be made in addressing the question ‘why should I be moral?’ if we began our theorising from the first-person standpoint.

The first stage of Wilson’s first-personal method is relatively uncontroversial. During the first eight steps of her argument, Wilson lays out clearly from her own first-person perspective how she thinks about morality. Her starting point is the idea that she is often judgmental and she has, by nature, preferences over most things. She also observes how evaluation permeates much of her thinking and language – almost any way in which she can describe an object entails something about how good it is.

Wilson then considers how plausible it is to think that some things are better for her than others. It is certainly better for her to go on existing right now than it is, for

example, to swim in shark-infested waters. Other things being equal, those very same things also seem to be better for others too. After this, she is led to consider how she can furthermore evaluate different forms of interaction in terms of whether they follow the ‘norms of civility’ that facilitate social interaction. Smiling when being introduced to someone, for example, is better than eating with mouth open and talking at the same time. Yet, these norms also raise new questions in cases in which what is good for me conflicts with the demands of the norms: for instance, why should I follow the social norms when it requires sacrificing my own interests? Here Wilson makes two poignant observations. Firstly, part of why in many cases she is inclined to follow the norms is the way in which she has been in the past rewarded for compliance. Likewise, she finds it pleasant to see her life run smoothly as a consequence of obeying the norms of civility. Yet, she also notes how she is able to opt out in some cases for good reasons and often no punishment follows.

At the next, eighth step, Wilson finally gets to her moral thinking. She begins by considering the content of the corresponding norms of morality. Wilson first concludes that morally wrong actions often involve obtaining a benefit to oneself whilst at the same time imposing a severe burden on others whereas morally right actions tend to involve sacrificing one’s own time and trouble in order to provide a benefit to others. In this way, moral norms are just like the norms of civility. Yet, there are also two differences: in the case of moral norms, the relevant harms to others tend to be more serious and also the moral norms are more universal than the norms of civility.

So far, Wilson has only described her moral thinking from the first-personal perspective. The relevant metaethical question then is: are the norms of morality recognised from the first-personal perspective real? Wilson's methodological claim is that thinking about this further question too from the first-personal perspective is fruitful. Her claim is that we can make progress in thinking about this question by considering the lessons we learnt from concerning first-personally why we should follow the norms of civility.

If I do not follow the norms of morality and others spot this, there will be often outrage, resentment and even punishments. Likewise, given how ingrained acting morally is in us, we similarly get pleasure (in the form of clear conscience) and other rewards from acting morally. Yet, from Wilson's first-personal perspective, these recognisable reasons do not seem to exhaust her reasons for following the moral norms. The further reasons have nothing to do with carrots and sticks but rather they seem to be grounded in our shared human way of life that cannot be captured from an external third-personal perspective. As Wilson points out, it is not clear whether from such a perspective there would always be reasons to be moral. Wilson thus makes a compelling argument to the conclusion that there is metaethical understanding that can only be gained through first-personal reflection.

The second genuinely new method of ethics is outlined in Andrew Sepielli's 'Consequentialism and the Evaluation of Action *qua* Action' (Chapter 6). Sepielli begins from the observation that there are many different standards for evaluating actions that range from thinking about which actions are good moves in chess to considering what we ought to do overall. The crucial difference between these

standards is that some of them are more authoritative than others. The natural question then is: which standards are the most authoritative ones?

Sepielli makes a striking claim in response to this question. He believes that the less authoritative standards evaluate the actions qua narrower categories of evaluation whereas the more authoritative standards evaluate actions more generally qua actions. Thus, for example, the standards of chess are superseded by the standards of overall ought because the former standards take into account only the considerations related to the goal of winning in chess whereas the latter take into account this consideration and also many other considerations, which bear on the evaluated action given all the ends we can have.

In order to find the most authoritative standard for evaluating actions, we then need to look for a standard that supersedes all others. This standard is the one that is the most liberal in terms of which ends and other considerations that are taken into account in the evaluation of action. Sepielli then argues that the traditional option – the evaluation of actions in terms of what we ought to do overall cannot be the most authoritative standard for evaluating actions. This is because there is a more general *uber-standard*, which evaluates actions qua events generally and therefore supersedes the standard of the overall ought. After all, the former standard places fewer constraints on what is taken into account in the evaluation of action. The overall ought limits our focus to what makes an action good one as an action whereas the *uber-standard* for evaluating event generally takes this standard into account plus a further set of considerations that make any event good or bad irrespective of whether it is an action or not.

At this point, Sepielli's view merges into a new method for doing first-order normative ethics. In first-order normative ethics, what we are most interested in is what the most authoritative standard for evaluating actions tells us to do. This is because we think both that moral demands are over-riding in virtue of their authoritativeness and that morality must in some way be practical and action-guiding. Sepielli thus outlines a two-stage method, which according to him we should apply when doing first-order normative ethics. We should first consider which standard of evaluating actions is the most authoritative due to the fact that it supersedes all other standards by being a more encompassing standard. As we have just seen, if Sepielli is right, that standard is the general standard for evaluating actions qua events.

At the second stage, we should then consider what that uber-standard is like. What is the way in which we evaluate, for example, how bad different natural disasters (as events that are not actions) are? Here Sepielli argues that the most general standard of evaluating events in general is a form of agent-neutral consequentialism that does not fetishise agency by making the value of outcomes depend on how they are related to the agency of different agents. After all, that form of consequentialism captures the way in which we tend to evaluate events that are not actions done by human agents. Sepielli thus offers us a new way of understanding how we should do first-order normative ethics. We are to consider first which standard of evaluating actions is most authoritative and then which types of considerations are taken into account in the evaluation of actions by that standard. By applying this method, Sepielli himself is led to endorsing a version of agent-neutral consequentialism but, of course, at this point

there is room for arguing that the method, when properly applied, should lead us to some other first-order ethical view.

Part 3: Evaluations of Recent Methods

The three chapters of Part 3 all evaluate some of the new methods that have been recently used in ethics. Christopher Cowie's 'The Similarity Hypothesis in Metaethics' (Chapter 7) is a defence of the recent 'metanormative' method in metaethics. Its defenders no longer focus on morality but rather investigate reasons for actions, beliefs and other attitudes more generally. The metanormative method thus accepts the 'similarity hypothesis', the idea that all reasons are reasons in the same way. That is, according to this assumption, the only difference between reasons in different domains is that they count in favour of different attitudes. Some reasons count in favour of beliefs whereas others in favour of desires, but there is no difference between what it is for those considerations to be reasons for the attitudes they support. If this is right, then investigating the relevant qualities of epistemic reasons should, for example, also help us to understand moral reasons.

The critics of this methodological approach argue that the similarity hypothesis is false because beliefs have a constitutive aim that can explain what it is for a consideration to be an epistemic reason whereas the practical attitudes for which there are supposed to be practical reasons lack similar constitutive aims. Beliefs aim at being true whereas it is not clear what a corresponding aim for, say, a desire could be. If beliefs and desires are different in this crucial way, then the metanormative method seems to fail: the assumption that epistemic reasons and other reasons are alike would turn out to be mistaken.

The main aim of Cowie's chapter is to argue that even the best formulation of the previous challenge to the metanormativity method provided by Stephen Darwall (2003) fails. In response to the previous objection, it could be suggested that practical attitudes such as desires do have constitutive aims: they aim at the good or the desirable (that is, a desire is correct if and only if it is a desire for something good). Darwall has, however, argued that, even if the aim of beliefs (truth) can ground theoretical reasons, the aim of desires (good) cannot do the same because there is an important difference between the kinds of aims truth and goodness are. His claim is that truth is a *substantial* aim of beliefs and thereby able to ground theoretical reasons whereas goodness is merely a *formal* aim of desires and hence incapable of grounding practical reasons. He thus argues that the aim of truth bears on what we ought to believe in a more substantial sense – it makes beliefs responsible to an external standard whereas, in contrast, the aim of desires, goodness, fails to do so. This is because goodness does not substantially restrict what we ought to desire based on a standard that would be external to practical reasoning. Because of this difference, Darwall argues that the aim of desire cannot ground practical reasons.

Cowie's objection to Darwall is based on the claim that a key premise in Darwall's argument can be read in two ways neither of which enables Darwall to reach his conclusion. Darwall claims that the crucial difference between beliefs and desires is that, because the aim of desire is merely formal, thinking about what to desire is not responsible to an external standard in the same way as belief formation is responsible to the standard of truth. Cowie observes that this premise can be read either metaphysically or epistemically. Under the metaphysical reading, the premise states

that, because the constitutive aim of desire is merely formal, the facts about what you ought to desire are such as they are in virtue of you doing some practical reasoning. In contrast, under the epistemic reading, the previous premise claims that, because desires lack a substantive aim, we can identify what we ought to desire only by following the internal norms of practical reasoning.

Cowie's central objection is that the previous interpretations of the central premise are not a problem for the defenders of the metanormative method. The metaphysical reading fails simply because the mere fact that desires lack a substantive aim cannot entail that what is good and desirable metaphysically depends on the satisfaction of the norms of practical reasoning. The epistemic reading, in contrast, cannot ground the required contrast between epistemic reasons and practical reasons. This is because in the epistemic domain too it is plausible that, in order to identify what we ought to believe, we cannot rely on external standards but rather we must follow the norms internal to theoretical reason. As a consequence, Cowie concludes that the best objection to the metanormative method in metaethics presented so far fails and so we should continue pursuing the metanormative project.

In his chapter 'The That' (Chapter 8), James Lenman begins by describing two very different approaches to ethical theorising. We could call the first method either 'voluntarism' or the 'way of invention'. That we should use this method seems to follow from the fundamental Humean thesis according which our passions – desires, plans, cares and concerns – constitute the source of both normativity and moral requirements. If this is right, then it seems natural also to think that normative theorising too should begin from the previous attitudes. The appropriate method for

finding out what is right and wrong would, according to this line of thought, be to first discover what we care about. After all, the talk of moral requirements and normativity is, on this view, a way of giving voice to our most fundamental concerns.

Lenman calls the second method, which he finds from the works of Iris Murdoch, Charles Taylor, and Susan Hurley, 'the interpretative theory'. This method emphasises the importance of the rich conceptual framework of thick evaluative concepts, which we inherit through growing up in our linguistic communities. The key idea is that, through being embedded in this conceptual framework, we can evaluate things in the world but also see, interpret, and understand ourselves too. Furthermore, by understanding the world and ourselves through the conceptual scheme of thick concepts, we are also able to shape our cares and concerns. This process can, furthermore, lead to a better understanding and interpretation of the thick conceptual scheme itself, which will then again lead to new, improved interpretations of ourselves.

Hence, whereas the first method assumes that the legitimate starting point of our moral theorising is what we most fundamentally care about, the second method sees ethical theorising to consist of something like an interpretative hermeneutic cycle in which the thick evaluations shape our understanding of our world and ourselves, which shapes our cares and concerns, which shapes our thick evaluative conceptual scheme, etc. Thus, on this view, ethical theorising fundamentally consists of interpretation that does not have any Archimedean starting point.

Lenman's central argument then is that, even if the second interpretative method is a more plausible description of how ethical theorising proceeds, this is not a reason to abandon Humeanism in metaethics. Rather, in addition to being able to accept the interpretative theory, the Humeans can also at the same time tame some of the metaethical mysteries, which the defenders of the interpretative method have left unresolved. The central insight of the interpretative model is that normative thought does not begin now. Rather, we gradually become a part of a whole community that has deliberated before and through this upbringing many of our community's evaluative commitments become ingrained in us. The claim then is that the thick evaluative language has evolved exactly for the purpose of voicing and communicating these commitments. The thick concepts thus contain our shared practical understanding in the form of cares and concerns that make us members of the same community. As we saw above, these concepts enable us to reflect and deliberate together which will lead to shaping our concerns in new ways. Thus, on the level of epistemology, the interpretative view is correct whereas as at the level of metaethics the Humeans can give an account of the attitudes that ground the deep evaluations, or so Lenman argues.

The final chapter of this section is Jack Woods's 'Footing the Cost (Of Normative Subjectivism)' (Chapter 9). This chapter is a discussion of a method for evaluating metaethical theories, which was first described by Philippa Foot (1972). According to this method, when we evaluate theories that attempt to describe the nature and content of moral judgments, we should not think only of whether the theory in question is true. Rather, according to Foot, we should also consider whether understanding our own moral judgments in the way suggested by the evaluated metaethical theory would

prevent those judgments from satisfying their functional role. The suggestion thus is that, if accepting a metaethical view about the nature of our moral judgments would, for example, make us stop making moral judgments altogether, then that theory could not be correct.

Foot herself applied this method in her argument against different versions of metaethical subjectivism and relativism. According to her, if we all started to think that what is right and wrong depended in some way on our contingent attitudes (which could have been easily different), we might all stop caring about our own moral judgments. Because it would be dangerous not to make any moral judgments, Foot then thought that we should not accept subjectivism or relativism in metaethics.

Woods begins by accepting the basic assumption behind Foot's method. He grants that metaethical views should be able to capture and make sense of the functional role of moral judgments in a way that can support our judgments' ability to carry out their functional role rather than undermine it. Thus, in this respect, Woods's methodology in metaethics is inspired by Foot's work. Yet, despite this, Woods wants to resist Foot's objection to subjectivism – his aim is to argue that Foot's method itself does not lead to the conclusion that we should reject subjectivism.

Woods's argument against Foot proceeds in two stages. Firstly, he argues that we should try to capture the functional role of moral judgments more carefully than Foot did. In the end, the role of these judgments is to enable us to co-ordinate our choices and actions in a mutually beneficial way. Fulfilling this function is possible only if we can make moral judgments that apply also to other people's actions. Furthermore, for

the moral judgments to be able to play the role of co-ordinating our choices and actions, we must be able to disagree with others by making conflicting moral judgments. Otherwise, we could not co-ordinate our actions by solving these disagreements through rational debates in which we take other people's views about our reasons seriously. It is only against this background that the potential arbitrariness of our moral judgments becomes a problem for the subjectivists. If our judgments and attitudes genuinely were arbitrary, it would not be clear how such judgments and attitudes could be used to persuade others to change their views in a way that would lead to action co-ordination.

In the second stage of his argument, Woods tries to respond to the arbitrariness challenge and to show that the subjectivist understanding of our normative judgments can support the functional role, which these judgments play. Firstly, the arbitrariness concern is supported by the idea that we could easily have wildly different sets of cares and concerns. Woods argues against this by defending the idea that our moral outlooks are relatively robust. After all, the cares and concerns that constitute our moral outlooks are what make us who we are and so being able to imagine having a completely different moral outlook would require being able to adopt the standpoint of a completely different person from their first-personal perspective. This is not something we are able to do easily.

Woods also argues that, even if according to the subjectivist framework we must evaluate the actions of others from our own moral standpoint that the objects of our evaluation might not share, the resulting judgments can still play the action co-ordinating functional role of moral judgments. This is because, even if the moral

standpoints of those with whom we disagree are different from ours, it is very rare that they are completely different. In most disagreements, solving the disagreement through rational debate can begin from at least some shared moral cares and concerns. If such common ground exists, then it is possible to change one's view in a moral disagreement rationally on the basis of the other's person's proposal even if that proposal is understood in the subjectivist way as an evaluation made from the speaker's own moral standpoint. As a consequence, Woods claims that, even if the method described by Foot is sound, it gives us no reason to reject subjectivist and relativist views in metaethics.

Part 4: Metaethics and Normative Ethics

The last two chapters of this volume, by Pekka Väyrynen (Chapter 10) and Matthew Silverstein (Chapter 11), are some of the first systematic attempts to answer a certain central methodological question that has always intrigued ethicists. It concerns the relationship between normative ethics and metaethics. There are two views about how these two domains of ethics are related to one another. On one view these two sub-fields of the discipline are independent of one another, whereas according to the other view they are intertwined areas of investigation. Which one of these views is correct will then have important consequences for how we should do research in ethics. If the fields are independent, then metaethicists do not need to care about what the correct views are in normative ethics (and vice versa), whereas if the fields are intertwined then, for example, finding the correct metaethical view will require taking part in first-order normative debates about what is right and wrong.

One crucial methodological question in ethics thus is: can the central questions of first-order normative be answered without taking a stand on metaethical issues and likewise is the correct metaethical framework neutral between different first-order ethical theories? Or, are the fields intertwined in such a way that normative ethicists and metaethics have to be done together at the same time? Even if many ethicists have taken a stand on these important methodological questions on the basis of their prior commitments in both normative ethics and metaethics, very few philosophers have considered these questions systematically. This is why the chapters by Väyrynen and Silverstein chapters are so interesting. They also answer the previous questions in very different ways.

In his chapter ‘Normative Commitments in Metanormative Theory’ (Chapter 10), Väyrynen argues that whether we should accept a given metanormative theory always turns on how plausible certain first-order normative assumptions are. If this is right, then metaethical inquiry will always also require doing first-order normative theorising. Väyrynen’s argument to this conclusion begins from the idea that any metaethical view concerning the meaning of normative terms or the nature of normative judgments and properties will end up referring to facts that make a normative difference according to that view itself. He then argues that any of the previous kind of considerations can make a normative difference only if they are normatively relevant. From this Väyrynen finally concludes that metaethical theories will always commit themselves to first-order normative claims about which features are normatively relevant.

Väyrynen also carefully illustrates how the previous argument applies to many well-known metaethical views. For example, consider the synthetic forms of naturalism according to which the reference of normative terms is determined by to which natural properties these terms are linked with the appropriate kind of causal connections. If we accept this view, then we must accept the idea that which natural properties causally regulate the use of our normative concepts must make a normative difference. After all, here facts about causal regulation will affect which actions are right and wrong. As a consequence, this view entails that facts about causal regulation must be normatively relevant, which itself is a normative assumption requiring first-order normative justification at the level of normative ethics. As this example illustrates, Väyrynen's argument has significant consequences for how metaethics should be done. If the argument is sound, metaethical theorising will always at least in part consist of considering first-order normative and moral questions.

Matthew Silverstein's chapter 'Revisionist Metaethics' (Chapter 11), in contrast, argues that in metaethics we should not be too concerned about the question of whether the first-order ethical implications of our metaethical views are intuitively acceptable. Silverstein first observes that different sides in the central metaethical debates tend to accept very different kind methodological principles. He calls those who value the explanatory power of metaethical views 'explainers' and those who value reflective equilibrium with our carefully considered first-order moral convictions 'preservationists'. Even if, of course, we would ultimately prefer a metaethical framework that would do well both explanatorily and extensionally, Silverstein sees the previous two methodological starting points to be pulling in different directions. Explanatory theories tend to be reductive and make normative

properties to depend on our attitudes, whereas extensionally intuitive theories tend to be non-reductive and thus able to recognise the kind of external and categorical reasons that match our intuitions.

Silverstein then argues that the explainers have a more appropriate methodology than the preservationists. This is because he believes that, given that we lack a sufficiently firm understanding of the subject matter of ethics, we should not care too much about the extensional adequacy of our theories. Silverstein then supports this main line of reasoning in two ways.

Firstly, Silverstein agrees with Alasdair MacIntyre (1981) and G.E.M. Anscombe (1958) that our moral vocabulary is historically based on a legalistic and theistic conception of ethics, which tends to understand right and wrong in terms of God's commands. Since we have abandoned the religious moral framework, we have, of course, made attempts to restructure our conceptual framework in a way that would not depend on God. Silverstein then suggests that, even if we should do moral philosophy in this new situation in which we are in the midst of revising our conceptual framework, we should still be sceptical about those elements of our moral practice that bear the closest traces of its religious and legalistic origins. One such element is the common thought that our moral judgments are universal – the intuition that these judgments apply to all agents independently of what they care about.

Silverstein's second argument begins from the idea that reasons must be essentially connected to practical reasoning. After all, reasons are considerations that bear on successful reasoning at least in some way. According to Silverstein, this means that

we should trust our intuitions about reasons only insofar as we understand the process of practical reasoning and what its constitutive aim is. He, however, thinks that we do not really understand practical reasoning. We know that its formal aim is to tell us what to do, but unfortunately, at the moment, we do not really know what the more substantial goals of practical reasoning would be. This is why Silverstein thinks that we should accept that at least some of our intuitions concerning what reasons we have will be wide off the mark. He also believes that, as a consequence, we should not rely on our substantial first-order intuitions about what reasons in evaluating those metaethical theories that have a lot of explanatory power. We should not be preservationists but rather explainers.

References

- Alfano, Mark and Loeb, Don (2017) 'Experimental Moral Philosophy', *The Stanford Encyclopedia of Philosophy*, Fall 2017 Edition, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2017/entries/experimental-moral/>.
- Anscombe, G.E.M. (1958) 'Modern Moral Philosophy', *Philosophy* 33, pp. 1–33.
- Brandt, Richard (1979) *A Theory of the Good and the Right* (Oxford: Clarendon Press).
- Daniels, Norman (2016) 'Reflective Equilibrium', *The Stanford Encyclopedia of Philosophy*, Winter 2016 Edition, Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/reflective-equilibrium/>.
- Darwall, Stephen (2003) 'Moore, Normativity, and Intrinsic Value', *Ethics* 113, pp. 468–89.
- Feyerabend, Paul (1975) *Against the Method* (London: Verso).

- Foot, Philippa (1972) 'Morality as a System of Hypothetical Imperatives',
Philosophical Review 81, pp. 305–16.
- Greene, Joshua (2013) *Moral Tribes: Emotion, Reason, and the Gap between Them and Us* (New York: Penguin).
- Hare, R.M. (1975) 'Rawls' Theory of Justice', in *Reading Rawls*, Norman Daniels. (ed.) (Oxford: Blackwell), pp. 81–107.
- Hooker, Brad (2000) *Ideal Code, Real World* (Oxford: Oxford University Press).
- Jackson, Frank (1998) *From Metaphysics to Ethics* (Oxford: Oxford University Press).
- Kelly, Daniel (2011) *Yuck! The Nature and Moral Significance of Disgust* (Cambridge, MA: MIT Press).
- MacIntyre, Alasdair (1981) *After Virtue* (Notre Dame: Notre Dame University Press).
- Mikhail, John (2011) *Elements of Moral Cognition: Rawls' Linguistic Analogy and the Cognitive Science of Moral and Legal Judgments* (Cambridge: Cambridge University Press).
- Rawls, John (1951) 'Outline for a Decision Procedure in Ethics', *Philosophical Review* 60, pp. 1–19.
- Rawls, John (1971) *A Theory of Justice* (Cambridge, MA: Harvard University Press).
- Rawls, John (1974) 'The Independence of Moral Theory', *Proceedings and Addresses of the American Philosophical Association* 47, pp. 5–22.
- Schwitzgebel, Eric and Cushman, Fiery (2012) 'Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers', *Mind and Language* 27, pp. 135–157.

Schwitzgebel, Eric and Cushman, Fiery (2015) 'Philosophers' Biased Judgments Persist Despite Training, Expertise, and Reflection', *Cognition* 141, pp. 127–137.

Singer, Peter (2005) 'Ethics and Intuitions', *Journal of Ethics* 9, pp. 331–352.

Sinnott-Armstrong, Walter (2008) 'Framing Moral Intuitions', In *Moral Psychology, Vol. 2, The Cognitive Science of Morality: Intuition and Diversity*, Walter Sinnott-Armstrong (ed.) (Cambridge, MA: MIT Press), pp. 47–106.

Williams, Bernard (1985) *Ethics and the Limits of Philosophy* (Cambridge, MA: Cambridge University Press).