INTRODUCTION

Jussi Suikkanen

Final Author Copy; published in J. Suikkanen and J. Cottingham (eds.): *Essays on Derek Parfit's* On What Matters (Wiley-Blackwell, 2009, pp. 1–20. Available online at https://onlinelibrary.wiley.com/doi/10.1002/9781444322880.ch1.

> The steady habit of correcting and complementing his own opinion by collating it with those of others, so far from causing doubt and hesitation in carrying it into practice, is the only stable foundation for a just reliance on it.[1]

There are few philosophers who have written two books that have given a new direction for the central debates in their discipline. Derek Parfit will soon join them. His 1984 book *Reasons and Persons* started new discussions on topics such as personal identity and population ethics.[2] His new treatise *On What Matters*[3] (soon to be published in two volumes) will likewise change the course of many fundamental debates in ethics.

Works like this are seldom created in isolation, but come to fruition through a long process of critical feedback. *On What Matters* is a prime example of this. For some years, successive drafts of the book have been circulating in the philosophical community. It has been exciting to see how the manuscript has developed as a result of the comments on these drafts.

As part of this process, the Philosophy Department at the University of Reading arranged a conference in November 2006 under the heading 'Parfit Meets Critics'.[4] This conference gave seven leading moral philosophers an opportunity to present their critical evaluations of *On What Matters* to Parfit in person. With one exception, all papers presented at this conference are included in the present volume. I hope that they will in part shape the future debates based on Parfit's work.

This introduction aims to present the main line of argument in *On What Matters*. I hope that this will help the reader to understand the nature and the significance of the critical reactions that follow. Providing a synopsis of the argument seems particularly necessary especially when it does not seem likely that Parfit's work will be published before this collection of essays appears. While explaining his views, I will also try to indicate which aspects of it will be attracting critical attention from the various contributors to the present volume.

It may be useful to say first something about the main motivation behind Parfit's overall view. He wants to defend the idea that there really are facts about

---

[1] See J. S. Mill, On Liberty (London: Routledge, 1991[1869]), p. 40.

[2] See Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984).

[3] This introduction (together with the essays that follow) will be based on the November 2007 draft of the manuscript. It was originally entitled *Climbing the Mountain*. All unattributed references are to this draft. Readers should be warned that the section-numbers of this draft may not always be the same as those in the eventual published volumes (for instance, Appendix A of the draft in question will be Part Six in the published books).

[4] I should like to thank Jonathan Dancy, Brad Hooker, Philip Stratton-Lake, Bart Streumer, and everyone else who took part in organising this successful conference. Thank are also due to all the speakers at the conference (i.e., who were the six contributors to this volume plus Jens Timmermann) and especially to Derek Parfit for his responses. Finally, I thank my co-editor John Cottingham for all his work on this volume, and him, Sven Nyholm, and Philip Stratton-Lake for comments on an earlier draft of this introduction.

which things matter in some fundamental sense, and the idea that morality is one of these things. These things that matter are not only what we happen to care about but also what we *ought* to care about. This basic conviction grounds two separate projects.

The first project is to argue that, when we say that we ought to care about something, we do not merely express our approval, make claims about what we want, or predict what would satisfy our desires. Instead, some things can themselves require that we care about them. The second project aims at showing that the best versions of the main moral theories in the end come to the same conclusions about what matters. Thus, there should be no major disagreements between rational inquirers such that they would undermine our conviction that some things really matter.

1. Reasons, Rationality, and Morality

Part One of Parfit's manuscript is largely preparatory, and yet the issues discussed here are significant in their own right. Parfit starts by taking the notion of reasons for desires to be a central normative notion in the practical domain (§1). Desires are states of being motivated for bringing about certain events (§3). If the event which is the object of a desire is desired for its own sake, the desire in question is a 'telic' desire. If the event is desired as a means, the desire in question is an instrumental one.

Some facts about the events that are the potential objects of the telic desires are reasons for having those very desires. Thus, I have a reason to want to eat salad if facts about eating salad (that it tastes good and is healthy) are reasons for wanting to do so – that is, they count in favour of desiring to eat it. Parfit argues that this 'being a reason for' relation between the facts about the desired objects and the desires for which they are reasons cannot be reduced to any other relations understood in broadly naturalist terms (App. A).

Besides reasons for desires, we also have reasons for actions. They are wholly derived from the abovementioned object-given reasons for desiring certain outcomes.[5] Thus, I have a reason for doing a given action, when doing that action is a way of bringing about an outcome which I have a reason to desire given by what this outcome itself would be like. All these reasons have strengths, and they can also add up and conflict. My reason to walk home (that it keeps me fit) can outweigh my reasons for taking the bus (that it is faster and easier [to take the bus]).

Because reasons are basic normatively speaking (that is, they cannot be accounted for by using any other normative or evaluative notions), one can try to use them to explicate other normative notions (§1–§2). On Parfit's view, other concepts are normative only if they entail claims about reasons. Thus, we use the term 'ought' in a normative sense only if we intend our claim 'x ought to φ' to entail that x has strong enough reasons to φ. Likewise, we use the term 'good' in a normative, reason-involving sense only if we intend our claim 'y is good' to entail that facts about y's nature give us reasons to react to y positively (by, say, admiring y). We can still say that something's being good, or what we ought to do, gives us reasons. But, if we do so, these reasons must derive all their normative force from the more basic good- and ought-making properties.

Parfit then introduces a distinction, which will play an important role (§2). This is the distinction between something's being good for someone and its being impersonally good. Parfit explains this distinction in the following way. Some of our

[5] Outcomes of actions are understood here in a broad sense. One outcome of an action is that the action itself is done. For this reason, facts about an action itself (or about its value) can be some of the reasons for desiring an outcome in which the action is done.

reasons are reasons for wanting certain events to take place for our own sake. These reasons (which needn't be selfish reasons) are self-interested reasons. They can be used to define what is good for us. Thus, some event is good for me when some facts about that event give me self-interested reasons for wanting the event to occur. The question of what is good for me is then the same substantial question as what events I have reasons for wanting to occur for my own sake.

As personal and partial reasons, self-interested reasons contrast with 'omnipersonal' reasons. These are the only reasons *all of us* would have for wanting events to occur if we considered them from an impartial point of view of a detached, uninvolved spectator. These reasons can be used to define what is impersonally good – that which we would have omnipersonal reasons for wanting to occur. If we have this kind of reasons to care equally about everyone's well-being, then everyone's well-being would be impersonally good. Parfit emphasises that these omnipersonal reasons can be recognised, weighed, and compared to our personal reasons from the same personal perspective from which we respond to our more personal reasons based on our self-interests and personal ties (§13).

At this point, Parfit's framework becomes more controversial. He first claims that the reason-providingness of facts does not usually depend on what we desire (§3–§9). Parfit calls this view of reasons 'value-based' (albeit it is the facts and not the evaluative properties that are reasons). In contrast, the views which claim that our reasons depend on our desires are 'desire-based' accounts of reasons. According to Parfit, these views hold that all reasons are provided by facts about what might fulfil our telic desires. On some of these accounts, our reasons are provided by what would satisfy our actual desires. According to other views, our reasons would be given by facts about what would satisfy the desires we would have if we were fully rational and informed.

For Parfit, the main problem with these views is that they cannot account for certain intuitive reasons (§7). It is conceivable that a fully informed person would have no desire to avoid a future of pure agony. On some desire-based views, this would entail that this person has no reason for not wanting such a future. This would be the case if having a desire to avoid future agony would not help to satisfy any other desires of the agent.

Some of these views require that our reasons are provided by facts about what would satisfy our fully *rational* desires. But, according to Parfit, these views can understand rationality here in only a purely procedural sense (ibid.). They can claim that our reasons are fixed by what we would desire if our desires were maximally coherent and consistent. However, if we make a substantially bad enough set of desires maximally coherent and consistent, we still cannot guarantee that the person will want to avoid a future of agony. If, in contrast, we stipulate that no-one counts as fully rational unless she desires to avoid future agony, then we are assuming substantial reasons instead of using rationality to account for what reasons we have. This view would no longer be a desire-based view.

Parfit's second controversial idea is that we can account for practical rationality in terms of responding to either real or apparent reasons (§10–§12). Many others think that practical rationality consists instead of some form of internal coherence – that our desires, intentions, plans, and evaluative and normative

judgments mesh with one another in the right way. Parfit argues against this view with counter-examples.[6]

He assumes that the defenders of this view are committed to thinking that, if I judge that I have reason not to care about the pain I might experience on future Tuesdays, then I am rational in not caring about my pain on Tuesdays. But, Parfit points out, such lack of concern would be foolish which is just what we mean when we say that someone is irrational. For this reason, the internal incoherence view of rationality must be mistaken.[7]

Parfit's alternative view is based, as already mentioned, on the idea of responding to reasons.[8] On this view, the rationality of a desire depends on its object. In desiring some event to occur, we have a variety of beliefs about the features of that possible event. The rationality of the desire then depends on whether these features (which we believe the object to have) would be good reasons for the desire.

Say that I want to run out of my office now because I believe that doing so would save me from the fire which I believe to be taking place. Because the latter belief is about a fact which, if it obtained, would be a reason for me to run out, my desire to run out is a rational one. It isn't foolish even if it turned out that my belief was irrational and false, and this house was not on fire. Even then I would be responding to an apparent reason. Similarly, if this house is on fire unbeknownst to me, I am not failing to respond to a reason even if I don't desire to run out, because I am unaware of the fact that would be a reason.

Thus, a desire is rational if the beliefs on the basis of which it is formed are about facts which would be reasons if they obtained. Notice that this account does not require that we have explicit normative beliefs about those facts being reasons. In the case that we have such normative beliefs, the corresponding desires will be rational only if the normative judgments themselves are rational in the sense of being supported by good reasons.

Parfit finishes off the first part with a discussion about how reasons relate to morality (§13–§16). He begins from Sidgwickian dualism about reasons.[9] According to this view, we cannot compare our reasons for bringing about what is best for us to our reasons for bringing about what is impersonally best. As a result, acting on either set of reasons would be rational. Parfit claims that this is wrong because in some cases the self-interested reasons really are weaker than the omnipersonal reasons. And, as already mentioned, our personal points of views allow us to compare our self-interested and other personal reasons to the omnipersonal reasons which we also have from these perspectives.

Given that our perspectives allow us to compare all kinds of reasons and all normativity is based on reasons, we can then formulate the profoundest problem of

---

[6] Parfit also argues against views according to which the rationality of a desire consists either of the desire being instrumental to the satisfaction of one's other desires, or of the desire being formed in a process of good practical deliberation and being stable under informed reflection (§12–§13).

[7] One problem with this objection is that the defenders of this view think that we should assess the rationality of a desire by looking holistically at how well it coheres with the agent's psychology overall (see Nomy Arpaly, *Unprincipled Virtue* (Oxford: Oxford University Press, 2003), ch. 2). In this case, a desire which is supported by one judgment about reasons can still be irrational if it fails to fit other judgments and concerns of the agent.

[8] At times, Parfit threatens to make this view true by definition. He writes that 'When I claim that such an act would not be rational, … I mean that, if we acted in this way, we would be seriously at fault for failing to respond to decisive reasons' (§13). This would make his view a concealed tautology.

[9] See Henry Sidgwick, *The Methods of Ethics* (London: MacMillan, 1874), 473.

morality in terms of reasons (§14). This is the question of whether we sometimes have sufficient reasons to act wrongly. The assumption here is that something matters only if we have sufficient reasons to care about it. We often assume that morality matters perhaps more than anything else. This would entail that we would always have sufficient reasons for being moral. Whether we have such reasons depends on what these reasons would be and how strong they are in comparison to all the other reasons. Parfit wants to show that the importance of morality can be vindicated by explaining what the overriding reasons for being moral are.

Of course, whether we have sufficiently strong reasons not to act wrongly depends on what we mean by 'wrong'. The last chapter of the first part explores this question to which, according to Parfit, there is no simple answer (§15–§16). By claiming that an agent did something wrong, we can mean a variety of things.

We can, for instance, use the term in fact-relative, belief-relative, and evidence-relative senses. When we say that an act is wrong in the fact-relative sense, we mean that the morally relevant facts count sufficiently against the action. In contrast, in the belief-relative sense, an act's wrongness depends on whether the morally relevant facts which the agent believes to obtain would count sufficiently against the action in the case that they really did obtain. When we discuss this kind of wrongness, what seems to matter is how probable we think the relevant possible outcomes of the actions are and how good they would be.

Even if the latter, belief-relative sense of wrongness is what we usually discuss when we talk about the actions we should not be doing, Parfit insists that we should also be interested in the fact-relative sense of wrongness. This is because fact-relative wrongness corresponds to which actions would be wrong in the belief-relative sense if we knew all the relevant facts. Only on the basis of this idea, we can try to figure out which actions are wrong relative to our beliefs when we don't know all the facts.

'Wrong' has also a variety of other senses according to Parfit. Sometimes we just mean indefinable 'mustn't-be-doneness' when we use this term. Other times we try to communicate that reactive attitudes like blame would be appropriate towards the action, or that the act would be unjustifiable or violate some important standards of conduct which we should accept. Parfit argues that we should not try to find any one, proper meaning of the term but rather endorse the rich ways in which the term can be used.[10] Parfit explicitly admits that, in asking whether we have sufficient reasons not to act wrongly, he will use the term in a loose, unspecified combination of the previously mentioned senses.

Two articles in this volume challenge the ideas just introduced. James Lenman opposes the idea that our talk about reason cannot be made sense of from a naturalist perspective by considering what we care about (Ch. 2 below). On Lenman's expressivist view, we talk about reasons in order to express our stable and widely held cares and concerns. The function of this talk is to enable us to search even wider unity and agreement in our community.

[10] Sometimes Parfit seems to suggest that what we mean by calling some action 'wrong' depends on in which sense we intend to use the term. This seems problematic. Wittgenstein convincingly argued that what we mean must instead be fixed by how other competent speakers would understand our claim in the situation irrespective of our intentions to mean something else (see Ludwig Wittgenstein, *Philosophical Investigations* (Oxford: Blackwell, 1953)). Furthermore, it is not clear that they could pick out what we meant by 'wrong' if the term really had as many meanings as Parfit suggests.

Parfit's objection to this expressivist view is that it leaves us without any reasons that would support the fundamental concerns which we would allegedly be expressing by talking about reasons. According to him, this would entail that nothing would really matter. To resist this, Lenman argues that our deep concerns form a network. Within the framework of this network, we can always be expressing some of our other deeply held concerns when we offer reasons for a given contested concern. This would mean that no concern would be held without a reason. We can also be expressing these very same actual concerns when we say that we would have a reason to care about something even if we didn't happen to care about it ourselves.[11]

Michael Smith challenges the idea that theories about reasons can be classified into the two mutually exclusive categories of value-based and desire-based views (Ch. 5). As noted above, some desire-based views understand reasons in terms of what would satisfy the desires we would have after rational deliberation and awareness of the relevant facts. Parfit claimed that on these views 'rational deliberation' can here mean only *procedurally* rational deliberation. The alleged problem then is that different agents can follow the purely formal principles of procedural rationality and end up with very different 'rational' desires (some of which would be foolish).

Smith sees a variety of problems in this view. In the case of some proposed principles of rationality it is very difficult to tell whether they are merely procedural or substantial. Some very procedural norms will furthermore commit agents to more substantial norms if these agents reflect on which desires it would be rational for them to have. And, some procedural norms (like the Kantian norm that requires us to have desires the satisfaction of which is consistent with others satisfying similar desires successfully) are often taken to have substantive consequences about what it would be rational to desire.

All of this would mean that substantive requirements of rationality might be derivable from merely procedural norms. Smith suggests that as a result we should accept a richer classification of views about reasons in terms of the kinds of principles of rationality they take as basic in accounting for reasons. On this view, the 'more desire-based' views would take only a few weak principles as basic and try to derive others from them. As a consequence, there would be some desire-based views which would have the same substantial consequences as Parfit's value-based view and the more controversial norms of rationality it takes as basic. Smith insists that the former views are better because they assume less.[12] He also worries that Parfit's framework gives us no explanation of how we could compare self-interested, personal, and omnipersonal reasons especially when Parfit matches these reasons with three distinct kinds of value without a generic notion of goodness that would encompass all three kinds.

2. Interpreting Kantian Ethics

In Part Two and the beginning of Part Three of *On What Matters*, Parfit turns to the question of whether we have always sufficient reasons not to act wrongly through examining Kant's views on ethics. Parfit's approach to Kant is original. Very roughly, the stereotypical view of Kant holds that he had one fundamental idea ('the supreme

[11] See Simon Blackburn, *Spreading the Word* (Oxford: Oxford University Press, 1984), 197–202.

[12] I wonder if there is a disagreement between Smith and Parfit which does not quite fit Smith's diagnosis. This disagreement seems to be about the order of explanation. On this proposal, desire-based views use norms of rationality to account for reasons, whereas value-based views use reasons to account for the norms of rationality. This difference would remain even if the views agree on what reasons there are and which desires are rational.

principle of morality') of what makes actions wrong and why we should always act morally. From this basic idea, we should be able to derive all the more specific principles of morality. He did, in addition, give us many arguments for this basic idea which he expressed in many ways in the different formulations of the categorical imperative.

Parfit's diagnosis of Kant is quite different. He first sets aside Kant's arguments for his basic idea as not very central.[13] He then claims that Kant's ethical writings are a rich source of many different ideas of the reasons why some acts are wrong and why we should not do them. On the basis of this assumption, he sets out to investigate how we should best understand all the interesting ideas Kant had. This is done by testing Kant's different ethical principles with possible counter-examples and seeing whether we could find formulations of them that would match our intuitions about these cases better. I will begin from Kant's ideas which Parfit thinks are indefensible and then move onto the ones which he thinks are better.

Parfit first rejects the Kantian idea that treating others as a mere means is something that makes one's actions wrong (§24–§26). He is more sympathetic to the idea that it is wrong *to regard* others as mere tools for one's ends. However, having this attitude is insufficient for making one's actions wrong.

To see this, we can begin from Parfit's best attempt to explain what it is to treat someone as a mere means (or to even come close of doing so). This would require (i) not being sufficiently guided by the concern for other's well-being, or (ii) not being willing to choose to bear a significant burden on behalf of the other person. But, now, take a gangster who visits the local cafeteria. Does he use the barista as a mere means for getting his cup of coffee? In the light of the previous criteria, this seems to be the case. After all, the gangster has no concern at all for the barista's well-being and he would bear no burden for him. Yet it seems unintuitive to think of this action as wrong even if perhaps the gangster's attitude is to be condemned.

Not even harming others as mere means is always wrong. Our gangster can save his child by injuring someone's toe (whom he considers as a mere means) without acting wrongly. Parfit argues that, no matter how we understand treating others as mere means, doing so is not always wrong whilst wrongdoing need not require treating anyone as a mere means.[14]

Another related Kantian idea is that we should treat all rational beings with respect (§27). Even if Parfit accepts that this is an important idea, he thinks that it fails to tell us what we should do.[15] On the one hand, we can understand the requirement of respecting others as a requirement not to treat them wrongly. But, in this case, the notion of respect could not be used to account for what it is to wrong someone on a threat of circularity. On the other hand, there are particularly disrespectful ways of treating others – of humiliating and ridiculing them. But, not all wrong actions are of this type. It is not even always wrong to act in these ways.

---

[13] These arguments are reconstructed and criticised in Appendix D.

[14] Parfit seems to agree here with Scanlon according to whom our intentions (to, say, treat someone as a mere means) cannot affect the wrongness of our actions even if, by giving meaning to our actions, they are relevant for assessing us as agents (see T.M. Scanlon, *Moral Dimension* (Cambridge, Ma.: Belknap Press, 2008).

[15] Stephen Darwall has recently argued that treating others with respect requires accepting that their claims and demands directly constrain our wills. He also argues that we have second-personal reasons to treat others in this way and that substantial conclusions about actions do follow from this (see Stephen Darwall, *The Second-Person Standpoint* (Cambridge, Ma.: Harvard University Press, 2006), esp. ch. 6).

Kant's related idea is that humanity in all rational persons has (as 'an end in itself') supreme, absolute value called 'dignity' which is always to be protected (§28–§29). Parfit accepts that, correctly understood, this idea is a profound truth. However, there are ways in which the claim is often misunderstood. Firstly, we cannot understand dignity here as a form of goodness. Even the worst criminals have dignity, but they are certainly not good. Secondly, humanity is often understood here in terms of rational abilities. It is these abilities that have dignity and ought to be protected at all costs. This cannot be right. Making people unconscious during surgeries makes them less rational but we should still do so to save them from the pain.

A third idea which Parfit rejects is Kant's idea of the Greatest Good which would consist of everyone being virtuous and getting the happiness they deserve (§30–§32) . Because this state of affairs would be the best one, we should all strive for it by promoting universal virtue and deserved happiness. Assuming that we cannot make others virtuous, the best way to promote this ideal in the current circumstances would be to make those who deserve happiness happy and those who deserve suffering suffer.

Parfit's objection to this view is that no-one ultimately deserves suffering for wrong-doing or happiness for acting rightly. We would deserve suffering and happiness in this way if we would have freely created our characters which determine how we decide to act. But, according to Parfit, such allegedly free acts of self-creation are not intelligible. For this reason, we should not make the wrong-doers suffer or the virtuous happy as Kant suggests.

At this point, the idea of universal laws creates a bridge to what are, on Parfit's view, Kant's more valuable contributions. This idea is often expressed by saying that it is wrong to act on a maxim that could not be a universal law (§33). This (inaccurate) formulation contains two instructive mistakes.

First, it takes the wrongness of actions to depend on the agent's maxims, i.e., on her policies for acting in certain ways (§35). This, however, is problematic. For instance, take the maxim 'Always do what is best for me!'. Presumably this maxim could not be universalised. Yet, that fact does not entail that an agent who always acts on this maxim acts always wrongly. Sometimes she will keep her promises (which is right), whereas other times she will break them (which is wrong).[16]

Second, the previous formula fails to condemn many bad maxims whilst threatening to rule out many good ones (§33). In principle, there is no reason why everyone could not decide to act on maxim 'Kill others!'. In this case, the previous formula would not condemn killing others. Maybe we could reply to this that surely some good people will not be capable of killing others. This might be right but it might equally well be that some bad people are causally incapable of helping others. The above formula would in this situation make helping others wrong.

To avoid the second problem, we should not care what maxims everyone could act on but rather concentrate on how one could rationally will them to act. This helps us to also avoid the first problem about the maxims. Thus, according to Parfit, the best formulation of the idea of universal laws or laws of nature is that *one acts wrongly if one does something one could not rationally will everyone to do in the similar circumstances* (§34–§35). Using this formula does not require knowing what

[16] It might be that we think less of the agent who keeps her promises on the egoistic grounds but this does not affect the permissibility of the action (see Scanlon, *Moral Dimensions*)

the agents' maxims are. It only requires being able to describe what they are intentionally doing and why.

When we apply the previous formula, we imagine what it would be like if everyone acted, whenever they could, in a certain way in certain circumstances.[17] We then ask whether we could rationally will the world to be like this. This is determined by whether we would have sufficient reasons to will that everyone acts in this way rather than in some other way.

This idea of being required to act in ways in which one could rationally will everyone to act is related to several other appealing Kantian ideas. First of all, Kant seemed to accept the consent principle according to which *it is wrong to treat others in ways to which they could not rationally consent* (§18–§19). Presumably people could rationally consent to some way of treatment when they could rationally will that everyone acts in this way in the similar circumstances (see §42). If this is right, then I should not only care about how I could rationally will everyone to act but also about how *everyone* could rationally will everyone to act in the given circumstances. By avoiding such actions, I can make sure that I don't treat others in ways to which they could not rationally consent.[18]

That there are such ways of acting requires that everyone shares a significant number of same strong reasons – that, for instance, burdens for some individuals are a reason for everyone not to will that everyone acts in the ways which cause these burdens. The previous consent principle thus requires that in most circumstances there is at least one way of acting to which everyone would have good enough reasons to consent.

Overall, however, this principle seems plausible. If someone has good enough reasons to refuse to give consent to some way of acting, these reasons are likely to provide good objections to that act. They should be stronger than the objections which others would have to the alternative ways of acting. We must also recognise that the consent principle only works if something like the value-based views about reasons and rationality are defensible. Otherwise there would not be anything which everyone could rationally will everyone to do. Furthermore, in some contexts actual consent counts in addition to the hypothetical rational consent, and the wrongness of some actions (like that of mistreatment of animals) has little to do with anyone's consent (§21–§22).

Parfit also claims that, if Kant would accept the previous natural law and the consent principles, he would not need to argue against the 'golden rule' in the way he did (§39). Roughly, according to this principle, we ought to treat others as we would want others to treat us. However, in its best form, this principle requires us to *treat everyone else as we would rationally choose to be treated if we were going to be in the position of everyone else and relevantly like them*. Even if this principle is based on a slightly different thought-experiment, it too refers to what everyone could rationally will. All these three principles also make the appealing impartialist assumption that everyone matters equally.

Finally, Parfit suggests that the appealing Kantian ideas about morality that are illustrated by the best versions of the consent principle, the natural law formula, and

---

[17] In order to use this formula, it is not enough to ask whether everyone could will everyone to act in some way when everyone else is acting in that way too. One must in addition ask can we also rationally will that everyone else acts in this way, whatever the number of people is who don't do that. Otherwise the view would be vulnerable to the so-called Threshold and Ideal World Objections (§36–§38).

[18] This amendment to the law of nature formula also helps with the so-called Rarity, High-Stakes, and Non-Reversability Objections (§40–§41).

the golden rule can be captured in a view which he calls Kantian contractualism (§45). On this view, *everyone ought to follow the principles whose universal acceptance everyone could rationality will*. Parfit suggests that this view is the best Kantian candidate for the supreme principle of morality which tells us what makes actions wrong and what are the strong reasons for not doing wrong actions. This principle also protects the vulnerable individuals who could be sacrificed for the small benefits of others. And, it has significant equalitarian consequences, because it recognises everyone's equal moral status and dignity as persons whose rational consent counts.[19]

Two articles in this volume challenge Parfit's views on Kantian ethics. Seiriol Morgan concentrates on defending an authentic Kantian view from Parfit's objections (Ch. 3).[20] Morgan argues that Parfit's conception of what it is to rationally will something conflicts with Kant's own views. Furthermore, with a correct understanding of this notion, there would be no need for making Parfit's radical changes to the Kantian framework. And, without such changes, Kantian ethics would not have any consequentialist consequences as Parfit claims it does (see below).

Morgan insists that we should return to the basic Kantian idea according to which whether one can rationally will that everyone acts in a certain way is determined by whether willing this would create a contradiction in one's will. After all, this is how Kant showed that it is not rational to will that everyone makes lying promises. Morgan argues that this conception of rational willing can explain why it is contradictory to will that no one helps others. The relevant contradiction would here consist of the agent willing something that would undermine her own freedom to which she is committed in virtue of having a will. Similar contradictions can be found also from the other cases which Parfit claimed to pose problems for the more basic formulas of the universal law.[21]

Gideon Rosen (Ch. 6) critically examines Parfit's candidate for the supreme principle of morality. As we have seen, according this principle, everyone ought to follow the principles whose universal acceptance everyone could will. Rosen has two main objections. According to Rosen's first objection, there are counter-examples to this principle. Assume there is a demon who would punish everyone if they accepted some valid moral principle. In these circumstances, one could not rationally will that everyone accepted this principle. As a result, not everyone could rationally will that the principle would be accepted by everyone. This would mean that it would be wrong for us to follow this principle, which certainly is counter-intuitive.

---

[19] Parfit recognises that this principle is close to Scanlon's contractualism according to which everyone ought to follow the principles which no-one could reasonably reject (see T.M. Scanlon, *What we Owe to Each Other* (Cambridge, Ma.: Belknap Press, 1998), ch. 4. These two principles both require that when we assess whether some principles are reasonably rejectable or rationally willable we cannot refer to antecedent views about whether these principles authorise wrong actions (§46). Parfit also argues that these principle are co-extensive because the principles which everyone could rationally will everyone to accept are the ones which no-one could reasonably reject, and vice versa (§55).

[20] Susan Wolf questions in the same spirit the authenticity of Parfit's interpretation of Kant (see Susan Wolf, 'Hiking the Range,' forthcoming in the second volume of Parfit's *On What Matters*).

[21] I wonder if there is more agreement between Morgan and Parfit than it might seem. Morgan claims that Parfit's view of rational willing is not Kantian because whether an individual could rationally will that everyone acts in some way would depend on how good the outcome would be for the agent (and this in part would depend on the agent's desires) (pages 00 and 00 below). But, this is not necessarily what Parfit had in mind. As noted earlier, his view takes rationality to depend on all the value-based reasons which the agent has (and not only on her self-interested reasons). These reasons might even include a reason to protect freedom in the Kantian sense.

According to Rosen's second objection, the proposed Kantian principle cannot be the supreme principle of morality as Parfit claims. A supreme principle of morality would tell us what ultimately makes actions right or wrong. On Rosen's view, such right- and wrong-making features must be much more basic than what the Kantian principle suggests. They would have to be first-order considerations such as that the act is kind, helps the neighbour, or causes pain.

3. Kantian Contractualism, Consequentialism, and the Master Argument

Let us assume that everyone ought to follow the principles whose universal acceptance (and the resulting reliable compliance) everyone could rationally will. What are these principles and why precisely should we follow them? The previous sections have not given us a satisfactory answer to these questions. Parfit's aim in the final chapters of Part Three of *On What Matters* is to provide them. Here Parfit presents an argument ('the master argument') to the rather surprising conclusion that the principles in question will be consequentialist principles.

Before this argument, Parfit explains how we should best understand consequentialism (§47–§48, see also §30–§31). Consequentialist views consist of two elements. The first, evaluative element specifies how we should assess the value of different outcomes. For Parfit, this depends on how good they are impersonally. According to Part One's definition of impersonal goodness, the impersonal value of an outcome is a function of how good reasons everyone would have for preferring that outcome from an impartial point of view. Thus, an outcome is impersonally best if everyone would have reason to prefer it from an impartial perspective.

The second, normative element specifies what we should do on the grounds that some states of affairs are evaluated as better than others. On a basic (impartialist) act-consequentalist view, this element says that we always ought to do the action which brings about as much impersonal value as possible. Thus, we should do what everyone would have most reason to want us to do from an impartial perspective. Parfit rejects this view because it would require wrong actions from us (§46).

In contrast, Parfit thinks that rule-consequentialism is a more plausible view. On this view, we first select the optimific principles. These are the principles that would make things go impersonally best if they were internalised by everyone or any other number of people. More simply, from an impartial point of view, everyone would have most reason to want these principles to be adopted. These reasons would be given by all the good things which the given principles would bring about (like general happiness, well-being, autonomy, biodiversity, knowledge, freedom from suffering, and the like).

Parfit is then finally in a position to put forward his master argument (§49–§55). It begins from the Kantian premise that everyone ought to follow the principles whose universal acceptance everyone could rationally will. Presumably anyone could rationally will whatever they would have sufficient reasons to will.

We can also assume that there is some set of principles such that its universal acceptance would make things go impersonally better than the acceptance of any other set (if we assume that acceptance always entails that the relevant principles will be reliable complied with). Given the earlier definition of impersonal goodness, this

optimific set of principles is the one whose universal acceptance everyone would have the strongest omnipersonal reasons to will.[22]

According to the next crucial premise, no one could have other conflicting (personal) reasons not to accept the optimific principles such that they would decisively outweigh the omnipersonal reasons to accept the optimific principles. Parfit argues that our self-interested reasons to accept other, non-burdensome principles are equalled by the altruistic, value-based reasons which we have for caring about the well-being of others (§50). The impartial perspective can also recognise the importance of our personal relationships (§51). For this reason, the optimific principles would leave us with room to act partially on our personal reasons given by our friends for instance.

From the previous three premises, it follows that everyone would have sufficient reasons to will that the optimific principles are adopted by everyone. According to the next premise, there are no other, non-optimific principles which everyone would also have sufficient reasons to accept. The idea behind this premise is that the acceptance of any other principles would always be bad for some people who would not therefore have sufficient reasons to accept them (§53).

The previous conclusion and the additional premise together entail that everyone would have sufficient reasons only to will that everyone adopts the optimific principles. Therefore, these rule-consequentialist principles are the only principles whose universal acceptance everyone could rationally will. And, because of this, they are the principles which everyone ought to follow. We have then argued from Kantian premises to a rule-consequentialist conclusion. Given that the optimific principles will probably resemble our current common-sense principles, it then seems that the major ethical theories can together vindicate many of our intuitive moral beliefs.

The three remaining articles in this volume contest different elements of the master argument. Michael Ridge worries about Parfit's way of selecting the optimific principles (Ch. 5). On Parfit's view, the optimific principles are the ones whose acceptance by any number of people would make things go best. According to Ridge, this understanding of rule-consequentialism is problematic because there may not be single principles which would make things go best on every level of acceptance. This would have the worrying consequence that nothing really is morally required according to Parfit's framework.

Ridge claims that there is a better way of selecting the optimific principles. On his proposal, the optimific principles make things go impersonally best on average if we take into account all possible levels of acceptance. Thus, we first assess how good are the consequences the principles would have if everyone accepted them. We then assess how good the consequences would be if 99% accepted them, if 98% accepted them, and so on all the way to the 1% acceptance-rate. At this point, we can finally count how good the average consequences of the different principles would be and pick out the principles that would have the best consequences on average. Ridge ends his article by being sceptical about the idea that everyone could rationally will to accept the optimific principles when they are understood according to his proposal.

Michael Otsuka argues that everyone has the strongest omnipersonal reasons to choose some non-optimific principles instead of the optimific ones. These

[22] This set would not arguably consist of the act-consequentialist principle according to which one ought to do whatever makes things go best. Accepting this principle would not have the best consequences because it would have harmful consequences for instance to our personal relationships (§54).

principles would ground non-consequentialist constraints against doing certain types of actions (Ch. 4). The initial problem with this claim is that Parfit seems to place no substantial constraints on which principles would be impersonally best, i.e., the ones which everyone would have most reason to accept from the impartial perspective. In principle, everyone could have most reason to accept principles which include constraints against, say, killing one person to prevent her from killing many others. As a result, the optimific principles would be non-consequentialist in the more traditional sense of the term.

Otsuka poses a dilemma for this line of thought. If we understand being impersonally best in terms of the reasons which everyone would have from the impartial perspective as suggested above, then Parfit's claim that everyone has strongest omnipersonal reasons to accept the optimific principles becomes trivial. This is because his version of rule-consequentialism lacks content until we specify what everyone would have reasons to accept. This would make the derivation of consequentialism from Kantian premises less interesting.

In order to resist this problem, the optimific principles could be selected in a more traditional way. On this understanding, these principles would create an outcome which has the most of the sort of value we should try to promote (for instance, well-being). However, according Otsuka, no one would have sufficient omnipersonal reasons to accept the optimific principles if we understand them in this way. In this situation, the relevant principles would require sacrificing individuals for the sake of the many. But, it could be argued, the moral status of individuals (which ought to be respected) gives us reasons not to accept principles which allow doing so.

Otsuka also argues that, contrary to what Parfit claims, Kantians and contractualists must be allowed to use their moral beliefs about rightness and wrongness when they assess which principles everyone could rationally will to be accepted. In pursuing reflective equilibrium, this must be allowed as long as rightness and wrongness themselves are not taken to be reasons for willing the acceptance of principles.

Jacob Ross argues that there is no sound argument from Kantian premises to the rule-consequentialist conclusion because two premises of the argument are false (Ch. 8). Ross first points out that in principle there could be many principles that are equally best from the impartial perspective. Because of our personal reasons, it could then be that different individuals would have most reason to accept different optimific principles. In this case, there would not be unique optimific principles which everyone would have most reason to accept as Parfit assumes.

Ross also argues that the optimific principles are not the only principles whose universal acceptance everyone is rationally permitted to will. He first claims that there could be non-optimific principles which everyone could rationally accept because no one would be harmed as a result of their adoption. This is because the only way in which these principles bring about less impersonal value is that they fail to bring about some public goods such as biodiversity.

Ross also thinks that there are some non-optimific principles whose global acceptance everyone could rationally will because every existing individual would benefit from their adoption. The impersonally best principles could require us to use a significant amount of our resources for improving the situation of the generations in a distant future. However, each of us would have more reasons to benefit ourselves and the people to whom we have close personal ties. If we acted on these reasons, then some other people would exist in the future than the possible people who would have existed if we had followed the optimific principles. In this case, everyone who ever

exists would have sufficient reasons to will the adoption of the non-optimific principles. This would mean that the Kantian premises would not lead to a requirement to accept the optimific principles.

I should like to end this introduction by expressing the hope that the problems raised by the articles in this volume will prompt others to follow Parfit in trying to find ways in which we could rationally agree about the things that really matter.

Jussi Suikkanen
University of Leeds
9 February 2009