

# *Parfit's final arguments in normative ethics*

Book or Report Section

Accepted Version

Hooker, B. (2021) Parfit's final arguments in normative ethics. In: McMahan, J., Campbell, T., Goodrich, J. and Ramakrishnan, K. (eds.) Principles and Persons: The Legacy of Derek Parfit. Oxford University Press, Oxford, viii, 479 pages. ISBN 9780192893994 doi: <https://doi.org/10.1093/oso/9780192893994.003.0010>  
Available at <https://centaur.reading.ac.uk/89642/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

Published version at: <https://oxford.universitypressscholarship.com/view/10.1093/oso/9780192893994.001.0001/oso-9780192893994-chapter-10>

To link to this article DOI: <http://dx.doi.org/10.1093/oso/9780192893994.003.0010>

Publisher: Oxford University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

[www.reading.ac.uk/centaur](http://www.reading.ac.uk/centaur)

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online



## Parfit's Final Arguments in Normative Ethics

Forthcoming in *Essays in Honour of Derek Parfit, Vol. 1: Normative Ethics and Personal Identity*, edited by Jeff McMahan, Tim Campbell, James Goodrich, and Ketan Ramakrishnan (OUP, 2020).

One of the most influential books in normative ethics has been Derek Parfit's *Reasons and Persons*. The moral theory dominating that book was Act Consequentialism. The first two volumes of Parfit's *On What Matters*, published twenty-seven years after *Reasons and Persons*, gave Act Consequentialism only passing attention. However, volume 3 of *On What Matters*, published six years later and in the week of Parfit's death, contains a long discussion that develops Act Consequentialism in ways that help it circumvent familiar objections. Yet then Parfit argues that Rule Consequentialism is a more plausible moral theory. In the sections that follow, I will provide more detail about the extent to which Parfit's normative ethics changed, I will explain how the reflective-equilibrium methodology he endorsed in the earlier volumes of *On What Matters* can lead to conclusions he did not there consider, and I will assess the kind of Act Consequentialism he discusses in the final volume of *On What Matters*. Finally, I will consider the arguments for Rule Consequentialism Parfit puts forward near the end of his final book.

### Section 1: *Reasons and Persons's* Act Consequentialism

The Act Consequentialism in *Reasons and Persons* was a criterion of moral rightness, holding that acts are morally right if and only if and because they produce at least as good consequences as any alternative acts. Act Consequentialism was not taken in that book to be a decision procedure appropriate for people to use in everyday moral decision making. *Reasons and Persons* defended Act Consequentialism against the charge that better consequences would come not only from people's not trying to employ Act Consequentialism as their procedure for making moral decisions but also from most people's not believing in Act Consequentialism. Here Parfit was following his hero Henry Sidgwick.<sup>1</sup>

*Reasons and Persons* also contained influential discussions of what Sidgwick had called Common Sense Morality. Common Sense Morality has an act-consequentialist component. The act-consequentialist component in common-sense morality is (a) a *general duty to do good*. Common Sense Morality also contains many deontological components.

---

<sup>1</sup> Sidgwick, *Methods of Ethics*. See also R. E. Bales, 'Act-utilitarianism: Account of Right-making Characteristics or Decision-making Procedure?'

These are (b) *prohibitions* on physically harming others, stealing, promise-breaking, and lying, (c) *personal obligations* to close family, friends, and others with whom one has personal connections, and (d) a *permission* to give one's own good somewhat more or less weight in one's decision making than the good of others.<sup>2</sup> In *Reasons and Persons*, Parfit didn't dismiss Common Sense morality. Instead, he argued that we could better achieve some of its aims if we focused on what we could achieve *together*, rather than individually.

## **Section 2: Volume 1 of *On What Matters***

*Reasons and Persons* was followed 27 years later by *On What Matters*, volumes 1 and 2. *On What Matters*, vol. 1 argued for the Triple Theory. The Triple Theory holds that (1) a revised version of Kant's Universal Law Formula, (2) Scanlonian Contractualism, and (3) Rule Consequentialism converge on the same set of moral rules. Act Consequentialism receives some attention in *On What Matters*, vols. 1 & 2, but Parfit's criticisms of it cut very deep. Moreover, Act Consequentialism does not form part of Parfit's recommended Triple Theory.

The fundamental goal of *On What Matters*, according to Parfit, is to resolve our disagreements about ethics, to show that the best ethical theories converge, to reveal that in normative ethics we are 'climbing the same mountain from different sides', with some prospect of reaching the same summit, from which we will be able to share the same views. In order to achieve this goal, Parfit thought he had to persuade us of certain things about each of Kantianism, Rule Consequentialism, and Scanlonian Contractualism.

The moral theory on which Parfit spends the greatest attention in volume 1 of *On What Matters* is Kant's. The Universal Law Formula is, in Parfit's view, the most intriguing of Kant's various versions of the Categorical Imperative. Kant's Universal Law Formula of the Categorical Imperative tells each to act on maxims that he or she can will to be universal laws. Parfit suggested Kant meant to refer not merely to laws that *the agent* could will but instead to laws that '*everyone* could rationally will that everyone accept and follow' (*On What Matters*, vol. 1, p. 340). Whether or not that is what Kant meant, Parfit argued that Kant's Universal Law Formula should be revised to become the foundational principle stating that everyone ought to comply with the principles that *everyone* could rationally will that everyone accept. The moral theory that consists of this foundational principle is dubbed by Parfit *Kantian Contractualism*.

---

<sup>2</sup> For a sustained campaign to incorporate many different deontological elements into consequentialism, see Douglas Portmore's *Commonsense Consequentialism*.

Since Kantian Contractualism holds that what is morally required, morally optional, or morally prohibited depends on what principles everyone can rationally will that everyone accept, we must ask what can be rationally willed. Parfit's answer was that what one can rationally will depends on what normative *reasons* one has. He distinguished four kinds of normative reasons relevant to what an agent could rationally will. These are (1) self-interested reasons for the agent to will what is beneficial to herself, (2) 'partial' altruistic reasons for the agent to will what is beneficial to people personally connected to her (e.g., her friends and family), (3) impartial reasons, and (4) 'deontic' reasons for the agent to will whatever is morally required. Parfit explains that 'deontic reasons' should be the output of our moral reasoning, not the input, on danger of circularity. So, if deontic reasons cannot be appealed to in determining what everyone can rationally will that everyone accepts, then what can be appealed are only self-interested, 'partial' altruistic reasons, and impartial reasons. These three kinds of reasons are then to be weighed against one another in order to determine what principles everyone can rationally will that everyone accept.

As indicated, Kantian Contractualism consists of the foundational principle that everyone ought to comply with the principles that everyone can rationally will that everyone accept. This *foundational* principle refers to *further* principles, which are to be selected on the basis that everyone can rationally will that everyone accepts them. Confusion can arise from using the term 'principle' to refer both to the foundational principle and to the further principles selected in accordance with the foundational principle. To avoid this confusion, I will use the term 'rules' to refer to the further principles.

With that terminological point made, I return to explaining the contentions that comprise Parfit's Kantian Contractualism. Parfit contended that the impartial reasons to will the universal acceptance of the rules whose universal acceptance would make the world go impartially best are sufficient for everyone to rationally will universal acceptance of these rules. Can that contention be right? After all, to be weighed up are not only impartial reasons but also reasons of self-interest and 'partial altruism'. Rules that might be *impartially* advantageous might be disadvantageous *for a given agent and her loved ones*. Nevertheless, since what any given agent can will for acceptance by everyone (everywhere and evermore, *On What Matters*, vol. 1, p. 382) will have immense ramifications across time for others outside the circle of the agent and her loved ones, the impartial reasons are strong enough not to be outweighed by the combination of reasons of self-interest and partial altruism.<sup>3</sup>

---

<sup>3</sup> This line of thought is not the same as the line of thought in *Reasons and Persons* that, *ex ante*, in conditions

Parfit also contended that the *only* rules that everyone would have sufficient reason to will, and could therefore rationally will, are the rules whose universal acceptance is favoured by the strongest impartial reasons (*On What Matters*, vol. 1, pp. 379–88). So, according to Kantian Contractualism, these impartially favoured rules determine right and wrong.

Parfit's Kantian theory does *not* hold that acts are right or wrong depending on whether everyone has sufficient reason to will that everyone do or not do these *acts*. The 'sufficient reason' test is about rules, not directly about acts. Acts are morally required, morally optional, or morally prohibited depending on what the impartially favoured *rules* are.

In this respect, Kantian Contractualism is like Rule Consequentialism. The question for proponents of Rule Consequentialism is not whether an act would achieve what the strongest impartial reasons favour. Rather the question is whether the act is permitted by the rules whose widespread acceptance would have the best consequences impartially considered. Parfit, like others before him, refers to such rules as the *optimific rules*.<sup>4</sup>

One of the points Parfit makes about Rule Consequentialism, especially in volume 3 of *On What Matters* (p. 417), is that Rule Consequentialism should be formulated in terms of the acceptance of rules, rather than compliance with rules. Imperfect beings like us might be incapable of perfect compliance. Even if achieving perfect compliance with any particular rule would be possible given enough training, practise, and reinforcement, the psychological costs of getting people to the point where successful compliance with the rule is secured might be very substantial. The costs to the teachers and the students of securing acceptance of rules that are more complicated would be higher than securing the acceptance of rules that are less complicated. Likewise, rules that demand more self-sacrifice have higher 'teaching and internalisation costs' than less demanding rules. At some point, the higher costs involved in getting more complicated and more demanding rules accepted will outweigh the behavioural advantages of compliance with more complicated or more demanding rules.<sup>5</sup>

Just as Parfit contended both (a) that, if Kantianism and Rule Consequentialism are to be acceptable and to be incorporated into a favoured convergence then Kantianism has to take a particular form and (b) that Rule Consequentialism has to be framed in terms of acceptance,

---

of uncertainty, each person could expect those to whom she is specially related to do better if everyone regularly chooses what impartial reasons support rather than what reasons of partiality support. For related discussion, see Johann Frick, 'Contractualism and Social Risk'; Korbilian R ger, 'On *Ex Ante* Contractualism'.

<sup>4</sup> As Michael Zimmerman reminded me, C. D. Broad used the term 'optimific' and formulated some of the innovations incorporated into what Parfit called 'act-involving Act Consequentialism', on which see section 3 below. Broad's Cambridge lectures on ethics can be found in his *Ethics*, edited by C. Lewy.

<sup>5</sup> See my *Ideal Code, Real World*, pp. 76–80, 89–90, 96–99, 138–41 and Parfit's *On What Matters*, vol. 3, pp. 414, 417–18, 420–21.

he contended that Scanlonian Contractualism needs to be modified. Scanlonian Contractualism holds that an act is wrong if it would be ‘disallowed by any set of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced general agreement’.<sup>6</sup> Scanlon himself placed an ‘individualist restriction’ on what can count as a reasonable objection to a proposed principle. This restriction is that reasonable rejection must be based on the effects on a single individual, not on the effects on an aggregate of individuals. Parfit argued that the individualist restriction has implausible implications (*On What Matters*, vol. 2, pp. 191–212).

Separately, since Parfit’s Kantian Contractualism makes the justification of rules depend upon impartial reasons to favour greater aggregate net benefit over lesser aggregate net benefit, Scanlonian Contractualism will have to drop its rejection of aggregate considerations if it is to converge with Parfit’s *Kantian Contractualism* on the selection of moral rules. One thing Parfit’s Kantian Contractualism shares with Rule Consequentialism is the appeal to a kind of impartiality that gives full weight to aggregative considerations. And so, if Scanlonian Contractualism is to converge with *Rule Consequentialism* on the selection of moral rules, again Scanlonian Contractualism will have to drop its rejection of aggregate considerations.

Rather than trace through the arguments Parfit put forward for the convergence of his Kantian Contractualism, Rule Consequentialism and Scanlonian Contractualism, I want to step back and observe that woven into *On What Matters*, vols. 1 & 2 was the presumption that our methodology in normative moral theory is the search for reflective equilibrium between all our beliefs, including our intuitions about kinds of cases and our most general moral principles.<sup>7</sup> Commitment to such a methodology crops up in vol. 1 of *On What Matters*, pp. 185, 352, 367, 370, 401, 415. Parfit reiterates his commitment to the reflective equilibrium methodology in vol. 2, e.g., p. 154, and very emphatically in vol. 3, p. 433.

Now, if our methodology is the search for reflective equilibrium, then the variety of consequentialism that comes into view is definitely Rule Consequentialism, since it coheres far better with our intuitions about what kinds of acts are right or wrong than familiar varieties of Act Consequentialism do. This observation about Rule Consequentialism is hardly new. Roy Harrod’s 1936 paper makes the observation, as does James Urmson’s 1953 paper.<sup>8</sup> The observation has been made countless times since.

---

<sup>6</sup> Scanlon, *What We Owe to Each Other*, p. 153.

<sup>7</sup> I pointed this out in ‘Must Kantian Contractualism and Rule-consequentialism Converge?’, at pp. 51–52.

<sup>8</sup> Harrod, ‘Utilitarianism Revised’; Urmson, ‘On the Interpretation of the Moral Philosophy of J. S. Mill’.

But, from about the middle 1960's until the middle 1990's, most philosophers thought that there are objections to Rule Consequentialism that are utterly devastating. The most influential objection was that Rule Consequentialism faces a dilemma: it either collapses into extensional equivalence with Act Consequentialism or is incoherent. That objection does not rely on a commitment to seeking reflective equilibrium between our moral theory and our intuitions about what kinds of acts are right or wrong. The objection was that Rule Consequentialism's internal workings are fatally defective.

However, in the middle 1990's, arguments were developed that Rule Consequentialism can be formulated so as to avoid both collapse into extensional equivalence with Act Consequentialism and incoherence. These arguments were convincing enough to revive interest in Rule Consequentialism.<sup>9</sup> If Rule Consequentialism is guilty of neither collapse nor incoherence, then its fit with our intuitions about right and wrong might endear it to us.

The search for reflective equilibrium between our moral intuitions and our moral theories pushes us to favour whatever versions of our theories have implications that cohere with our intuitions. Thus, it can be no surprise that what seem to us the best versions of moral theories (such as the best versions of Kantianism Contractualism, Rule Consequentialism, and Scanlonian Contractualism) have very similar implications.<sup>10</sup> In light of that point, there is a missed opportunity in *On What Matters*, vols. 1 & 2, which I will now identify.

In the first two volumes, Parfit wrote little directly about Common Sense Morality, although he invokes common-sense moral intuitions to reject Act Consequentialism (vol. 1, pp. 144, 191, 362, 404–5, 417, 482) and to reject various Kantian theses. As I mentioned earlier, Common Sense Morality is a pluralistic theory consisting of (a) a *general duty to do good*, (b) *prohibitions* on physically harming others, stealing, promise-breaking, and lying, (c) *personal obligations* to close family, friends, and others with whom one has personal connections, and (d) a *permission* to give one's own good somewhat more or less weight in one's decision making than the good of others. Common Sense Morality consists of the same general and specific obligations, prohibitions, and permissions that the best versions of Kantian Contractualism, Rule Consequentialism, and Scanlonian Contractualism endorse. Hence, when in vols. 1 and 2 of *On What Matters* Parfit put forward his Triple Theory of Kantian

---

<sup>9</sup> I first addressed the objections in 'Rule-consequentialism, Incoherence, Fairness'. There I proposed that a rule-consequentialist agent's most basic moral motivation could be a concern for impartial defensibility rather than concern for the impartial good. See also my *Ideal Code, Real World*, ch. 4; and 'Rule-consequentialism and Internal Consistency'. More recent discussions of the topic include Susan Wolf's 'Two Concepts of Rule Utilitarianism' and David Copp's 'The Rule Worship and Idealization Objections Revisited and Resisted'.

<sup>10</sup> I made this point and the points below about extending this line of thought in 'Must Kantian Contractualism and Rule-consequentialism Converge?', at p. 52, including fn. 10.



Contractualism, Rule Consequentialism, and Scanlonian Contractualism, he should have included Common Sense Morality, thereby producing *the Quadruple Theory*.

Common Sense Morality is unlike the other three theories just mentioned in that Common Sense Morality, I take it, is *agnostic* on whether there is a deeper moral principle providing a foundational and unifying justification for the general and special duties, the prohibitions, and the permissions. Such agnosticism about a foundational unifying justificatory principle contrasts with the commitment to such a foundational unifying justificatory principle made respectively by Kantian Contractualism, Rule Consequentialism, and Scanlonian Contractualism. These three theories disagree with one another about what the one unifying foundational principle is. But these three theories' disagreement about what the unifying foundational principle is doesn't keep the three theories from being the elements of Parfit's Triple Theory. If disagreement at the foundational level doesn't preclude theories' being parts of the Triple theory, then a theory's agnosticism about whether there is such a correct unifying foundational principle cannot be a bar.<sup>11</sup> Parfit himself later, *in effect*, admits that Common Sense Morality should be included in the convergence (vol. 3, p. 434).

This line of thought might be extended even further. We should be open-minded, I think, that the best version of *Virtue Ethics* will have implications that converge with the prohibitions, duties, and permissions endorsed by the Quadruple Theory (cf. *On What Matters*, vol. 3, p. 418). If we find such a Virtue Ethics, then we have the *Quintuple Theory*—consisting of Parfit's Kantian Contractualism, Rule Consequentialism, Scanlonian Contractualism, Common Sense Morality, and Virtue Ethics.<sup>12</sup>

### **Section 3: Act Consequentialism in *On What Matters*, vol. 3**

Can Act Consequentialism be refined to a point where it can be included in the convergence? Having spent much of *Reasons and Persons* on Act Consequentialism but little of vols. 1 and 2 of *On What Matters* on this theory, Parfit devotes more of vol. 3 to discussing Act Consequentialism than he does to discussing any other normative moral theory. I find that surprising, though I guess that in epic journeys the protagonist often circles round past earlier sites.

---

<sup>11</sup> I distinguish between Common Sense Morality, which I take to be agnostic about a correct unifying foundational principle, and deontological pluralism, which, as its name suggests, denies there is a correct unifying foundational principle. But whether or not I am mistaken about Common Sense Morality's agnosticism, Parfit should have included Common Sense Morality in a Quadruple Theory.

<sup>12</sup> Does a commitment to moral realism militate in favour of, or against, believing that various normative ethical theories converge? See Marius Baumann, 'Parfit, Convergence, and Underdetermination'.

In *Reasons and Persons*, Parfit commented that Act Consequentialism can appeal to non-utilitarian principles. He wrote,

According to three such principles, it is bad if people are deceived, coerced, and betrayed. And some of these principles may refer to past events. Two such principles appeal to past entitlements, and to just deserts. ... If our moral theory contains such principles, we are not concerned only with *consequences* in the narrow sense: with what happens *after* we act (p. 26).

In *On What Matters*, vol. 3, Parfit expands on the extent to which Act Consequentialism's counterintuitive implications can be avoided if the theory postulates that *acts* can have 'intrinsic moral badness' or 'intrinsic moral value' (vol. 3, pp. 396–406). For example, if acts of promise-breaking have intrinsic moral disvalue, then we could explain *in Act-Consequentialist terms* how a particular act of promise-breaking can be morally wrong even if this act of promise-breaking benefits people slightly more than would the alternative act of promise-keeping. In this particular example, the intrinsic moral badness in the act of promise-breaking outweighs the slightly greater benefit to people that is produced by promise-breaking. Thus, here the act of promise-breaking would not be the available act with the greatest overall goodness.

The intrinsic moral badness here is, in Parfit's terminology, 'non-deontic', meaning that the badness is not wrongness. Parfit wants to allow that, though a kind of act has intrinsic moral badness, and this often will be part of why an instance of the act is morally wrong, there can be instances of the very same kind of act which are not morally wrong, though, like all other instances of the kind, still intrinsically morally bad. An example might be one in which an act of deceit or coercion or promise-breaking or even harming an innocent person, though intrinsically morally bad, would not be morally wrong because the intrinsic badness would be outweighed by tremendous instrumental value to something good. Suppose, for example, an intrinsically bad act of deceit or coercion or promise-breaking or even harming an innocent person were necessary to save very many innocent lives (vol. 3, pp. 398–401). In such an example, the act of deceit or coercion or promise-breaking or even harming an innocent person would be intrinsically bad but would not be morally wrong.

Parfit uses the term 'Act-involving Act Consequentialism' to refer to the theory that the rightness or wrongness of an act is determined exclusively by whether its consequences are at least as good as those of any alternative act, where the goodness of consequences includes not

only the impartial value or disvalue in states of affairs but also intrinsic moral value or intrinsic moral disvalue of acts. Acts with intrinsic moral value might be acts of gratitude or loyalty or treating people as they deserve (vol. 3, p. 406). Treating people as they deserve, for example, would be the right thing to do, according to Act-involving Act Consequentialism, even if the consequences would contain a little less aggregate well-being than not treating them as they deserve would. But in cases where treating people as they deserve would result in far less aggregate well-being, the Act-involving Act-Consequentialist calculation would come down in favour of not treating people as they deserve.

Act-involving Act Consequentialism offers an explanation that is different from but compatible with the deontological explanation of the wrongness of these acts offered by W. D. Ross in terms of *prima facie* duties. In both Act-involving Act Consequentialism and Ross's pluralist deontology, there are various moral properties of acts that always count with the same moral polarity whenever these properties are instantiated. For example, any act with the property of being an act of promise breaking *has some intrinsic moral disvalue* and *there is a prima facie duty weighing against it*. According to both Act-involving Act Consequentialism and Rossian deontology, the fact that an act has the property of being an act of promise breaking is always morally negative, but this negative doesn't always outweigh other morally relevant considerations (vol. 3, pp. 404–405).

Pointing out how Act-involving Act Consequentialism can operate in parallel fashion to Ross's theory of *prima facie* duties raises two possible objections to Act-Involving Act Consequentialism. The first of these objections to Act-involving Act Consequentialism Parfit addresses. The second he does not.

The first objection is that Act-involving Act Consequentialism seems inferior to Ross's theory of *prima facie* duties when a distinction is made between duties in general and duties owed to people with whom one has special connections.<sup>13</sup> Ross stressed this distinction. In contrast, the Act-involving Act Consequentialism that Parfit discusses focuses on the intrinsic goodness or badness of kinds of act and on consequences, *assessed agent-neutrally*.<sup>14</sup>

Here is a standard example. Suppose there is intrinsic moral goodness in people's helping their own children. There might thus be an agent-neutral duty to promote the extent to which people in general help their own children. Such a duty can be distinguished from an agent-relative duty that parents have to help their own children. In a situation in which somehow my

---

<sup>13</sup> For an influential discussion, see Ross, *The Right and the Good*, p. 25.

<sup>14</sup> In contrast, see Portmore, *Commonsense Consequentialism*.

helping my children will prevent other people helping their children, I am under pressure from the agent-relative duty to help my own children. This pressure might be opposed by a duty to promote the welfare of others in general and would be opposed by any agent-neutral duty to promote the extent to which parents in general help their children. Where there is such a conflict of duties, depending on what is at stake for the various effected parties, my duty to my own children might outweigh my agent-neutral duty to promote the extent to which people in general help their children, and outweigh even the duty to promote the welfare of others in general.<sup>15</sup> In other possible cases, the agent-relative duty to my own children is outweighed by agent-neutral duties.

Other examples of agent-relative duties zero in on the relation of agents to the acts they do, as opposed to acts they promote or minimize. We have an agent-neutral duty try to prevent the harming of innocent people, but we have a stronger agent-relative duty not to harm people. We have a duty to oppose the coercion of people, but we have a stronger duty not to coerce people ourselves. And so on.

Ross's theory of prima facie duties stresses agent-relative duties. This is also true of what I earlier referred to as Common Sense Morality. In contrast, Parfit's Act-involving Act Consequentialism is unsympathetic to agent-relative duties (vol. 3, 406–12).

He distinguishes between cases in which others are behaving wrongly and cases in which others are not behaving wrongly. He proposes that most of us, and perhaps even Common Sense Morality, would agree that, in cases where others are *not* behaving wrongly, our duty is to minimize harm, deceit, or coercion, rather than to avoid being the one who perpetrates the harm, deceit, or coercion. The idea is that, at least in these cases where others are not behaving wrongly, we should focus on minimising the amount of bad, rather than focus on avoiding doing anything with intrinsic moral badness in it. I am doubtful that most people will agree with him about that.

Parfit thinks cases in which others are behaving wrongly are different. Suppose we are threatened that, unless we kill some others, many more others will be killed by other agents. About such a case, Parfit writes, 'we may plausibly believe that it would be wrong to give in to such threats, since that would encourage future threats, in ways that might then lead to more people being killed.' (vol. 3, p. 412) This response remains completely agent-neutral Act Consequentialist. In so far as we think morality is at least partly agent-relative, the complete

---

<sup>15</sup> For an especially widely discussed attack on agent-relative duties, see Samuel Scheffler's *The Rejection of Consequentialism*.

agent-neutrality of this kind of Act-involving Act Consequentialism will strike us as counter-intuitive.

Even more importantly, the earlier example in which I was weighing up my agent-relative duty to help my own children against my agent-neutral duty to promote the extent to which people in general help their own children was not one in which my helping my children will somehow lead to other parents' *wrongly* failing to help their children. The situation might well be such that my helping my children will merely result in other people's not having the opportunity to help theirs. To the extent that our intuition remains that parents can have a duty to help their own children even when their helping their children would result in other people not having the opportunity to help theirs, Parfit's arguments against agent-relative duty are unpersuasive.

A different objection to Act-involving Act Consequentialism is utterly independent of the above arguments. Suppose Parfit were to respond to the above line of objection by conceding that Act-involving Act Consequentialism should incorporate agent-relative duties just as much as necessary in order to make the theory have implications that converge with our intuitions. Even if he responded in such a concessionary way, the objection I am about to offer would work just as well, perhaps even better.

This objection is that, just as Ross's pluralist deontology relies on a lot of postulates about prima facie duties, Act-Involving Act Consequentialism contains a lot of postulates about the intrinsic moral value or disvalue of acts. Admittedly, these postulates enable Ross's pluralist deontology and Act-Involving Act Consequentialism to have intuitively plausible implications about what is right or wrong in various circumstances. However, if some other moral theories are equally able to generate intuitively plausible results and these other theories do so on the basis of fewer postulates that are at least as attractive, then these other theories have more explanatory power. Parfit's Kantian Contractualism, Rule Consequentialism, Scanlonian Contractualism, and Virtue Ethics may or may not be otherwise defensible, but at least they offer (albeit different) unified explanations of why morality generally opposes harming others, stealing, coercion, promise breaking, failing to give special weight to one's family and friends, etc. As long as Parfit's Kantian Contractualism, Rule Consequentialism, Scanlonian Contractualism, and Virtue Ethics have implications that are either the same as or at least as intuitively attractive as the implications of Rossian pluralist deontology and Act-Involving Act Consequentialism, the greater explanatory power of Parfit's Kantian Contractualism, Rule Consequentialism, Scanlonian Contractualism, and Virtue Ethics gives

them important advantage over both Ross's pluralist deontology and Act-Involving Act Consequentialism.

Parfit did not end up endorsing Act-Involving Act Consequentialism. He accepted that it loses out to Rule Consequentialism. The next section puts his arguments for Rule Consequentialism under the spotlight.

#### **Section 4: The Arguments for Rule Consequentialism in *On What Matters*, vol. 3**

The end of *On What Matters* contains three arguments for Rule Consequentialism.

One of the arguments for Rule Consequentialism that Parfit explicitly makes is the familiar argument that Rule Consequentialism slots into reflective equilibrium with our other moral beliefs. Here is his most explicit statement of the argument:

We may start by accepting what seem to us the most plausible principles of Common Sense Morality. We then ask whether these principles can all be given some further justification, which may appeal to some feature that these principles have in common. On one plausible hypothesis, the best principles of Common Sense Morality are also the principles whose acceptance would on the whole make things go best. We might justifiably accept this hypothesis. These beliefs would then support a wider theory which combined this version of Common Sense Morality with Rule Consequentialist justification.

These two parts of this wider theory would achieve more by being combined. Rule Consequentialism would be strengthened if this theory supports what seems to be the best version of Common Sense Morality. The version of Common Sense Morality would be similarly strengthened if it can be plausibly supported in this Rule Consequentialist way. (*On What Matters*, vol. 3, p. 433. See also pp. 421–22, 450.)

A familiar line of objection to this argument is that, even if Rule Consequentialism has practical implications that are extensionally equivalent with Common Sense Morality's requirements and permissions, Rule consequentialism's *explanations* of these requirements and permissions strikes us as counterintuitive. This line of objection rejects Parfit's 'On one plausible hypothesis, the best principles of Common Sense Morality are also the principles whose acceptance would on the whole make things go best. We might justifiably accept this

hypothesis.’ I will not pause to consider the line of objection, since I have tried to answer it elsewhere.<sup>16</sup>

Parfit’s second argument for Rule Consequentialism begins with the observation that humanity now faces many challenges, and if we go on behaving as we have done, the cumulative effects will be terrible for others, especially future generations. In particular, Parfit is considering cases in which an individual’s contributing to global warming or polluting the environment is morally wrong but this one individual’s action has very small or even imperceptible effects on anyone else. Parfit explores the possibility that we can make progress with such problems by thinking in more careful Act-Consequentialist ways. In this discussion, Parfit is developing earlier lines of thought he explored in ch. 3 of *Reasons and Persons*, (pp. 67–86). In vol. 3 of *On What Matters*, however, when Parfit concludes this discussion of the wrongness of acts that harm very many but harm each to only a very small or even imperceptible degree, he admits that the explanation that points to the aggregate of many small harms caused by the individual’s act is inferior to an explanation that appeals, ‘not to the separate effects of particular acts, but to the combined effects of what we and others together do. Some act would be wrong, we believe, if all optimific rules would condemn such acts.’ (p. 432)

Parfit continues,

[T]he wrongness of these acts can be best explained in this Rule Consequentialist way. We are considering acts that would be very slightly worse for each of very many people. ... Imperceptible amounts of pain, and other such harms, seem to most of us to be below any plausible threshold of moral significance. If we are Rule Consequentialists, however, we deny that each of these acts is made to be wrong by this act’s effects. These acts are wrong, we believe, when and because they are condemned by optimific rules. Whether some rule is optimific depends on whether things would on the whole go better if most of us, or many of us, accepted and tried to follow this rule. Because these claims are not about the effects of single acts they are not challenged by the fact that, in the cases we have been considering, no single act would have perceptibly bad effects. (p. 432)

---

<sup>16</sup> See my ‘Ross-style Pluralism versus Rule-consequentialism; *Ideal Code, Real World*, ch. 1; and ‘Reflective Equilibrium and Rule Consequentialism’.

This argument that the wrongness of acts that are very slightly worse for each of very many people is better explained by Rule Consequentialism than by Act Consequentialism was foreshadowed by Parfit's remark in *Reasons and Persons*, 'These are the cases where we naturally say, "What if everyone did that?"' (*Reasons and Persons*, p. 66). And he pointed out in *Reasons and Persons* that most views about the nature of morality, including Kantianism and Rule Consequentialism, share the idea 'that morality is essentially a collective code' (*Reasons and Persons*, p. 106). If moral requirements, prohibitions, and permissions are part of a collective code, we should consider the consequences of collective acceptance of and compliance with this code. Furthermore, *Reasons and Persons* contained an arresting discussion of the attractions of 'Collective Consequentialism' (pp. 30-31). These attractions not only were taken up in work by Liam Murphy, Tim Mulgan, and me but also retained their appeal to Parfit to the end.<sup>17</sup>

I will conclude this section by pointing to a third argument for Rule Consequentialism that Parfit presents at the end of vol. 3 of *On What Matters*. This argument starts by considering which patterns of motives and dispositions would make the world go best. Naively, we might think that the world would go best if everyone were most keenly concerned to maximize the impartial good. However, as Parfit writes,

If most of us were pure Act Consequentialists who were most strongly motivated to do whatever would make things go best, our acts would have many good effects. But our lives would on the whole go better if most of us had some other strong motives and try to follow some other policies or rules. It is good, for example, that most of us strongly love our close relatives and some friends. Having such love and being loved are some of the greatest goods in most people's lives. If instead we cared equally about everyone's well-being, we would have no strong love for anyone, but only what Aristotle called 'watery kindness'. This would not be how things could go best. (*On What Matters*, vol 3, p. 420)<sup>18</sup>

The conclusion here is that, if the choice is limited to either (a) most people's accepting Act Consequentialism and having the motivation that matches it or (b) most people's accepting

---

<sup>17</sup> See, for example, Murphy, *Moral Demands in Nonideal Theory*; Mulgan, *Future People*; and my *Ideal Code, Real World*.

<sup>18</sup> Here Parfit is drawing on Sidgwick, p. 434. The point is one that Parfit has long accepted—see *Reasons and Persons*, pp. 27–8, 30. Worth comparing is Garrett Cullity's *The Moral Demands of Affluence*, ch. 7.



some improved version of Common Sense Morality along with the motivations it allows, then things would go better if most people accept some improved version of Common Sense Morality.

And, as we have seen, Parfit maintains that the motives and rules that some improved version of Common Sense Morality would endorse would also be the optimific motives and rules—that is the ones whose acceptance would make things go best. And so these would also be the motives and rules endorsed by Rule Consequentialism. Hence, Parfit’s succinct formulation of his third argument for Rule Consequentialism:

What matters most is how well things go.

Things would on the whole go best if we have optimific motives and we accept and tried to follow optimific rules.

Therefore,

We ought to have such motives and we ought to try to follow such rules. (p. 432)

Parfit goes on to consider a possible objection to the second premise of this argument. This premise considers only two possible scenarios. In the first scenario, most people accept Act Consequentialism and have the corresponding motivations. In the second scenario, most people accept Common Sense Morality and Rule Consequentialism and have the corresponding motivations. The objection is that there is a third scenario that should have been considered. This is a scenario in which there is a ‘mixed population’, a population in which *some* people accept some improved version of Common Sense Morality and Rule Consequentialism while *other* people are pure Act Consequentialists. Thinking about a mixed population is a thought experiment that was given its classic formulation over one-hundred and fifty years ago by Sidgwick.<sup>19</sup> It is also a thought experiment periodically posed in Parfit’s work (*Reasons and Persons*, p. 30; *On What Matters*, vol. 3, p. 414).

I do not mean to suggest that the only possible ‘mixed-world’ scenario worth considering is the one in which in which some people accept some improved version of Common Sense Morality and Rule Consequentialism while other people are pure Act Consequentialists. Which mixed-world scenarios are worth considering seems to me too large a question to be addressed in a paper focused on Parfit’s arguments. The only mixed-world

---

<sup>19</sup> Sidgwick, pp. 489–90.

scenario on which Parfit comments is the one in which some people accept some improved version of Common Sense Morality and Rule Consequentialism while other people are pure Act Consequentialists. And so this is the mixed-world scenario targeted here.

The justification Parfit offers late in vol. 3 of *On What Matters* for rejecting this scenario is as follows:

‘I am assuming that ... everyone ought to have the same moral beliefs.  
Moral truths are not true only for certain people.’ (p. 420)

The conjunction of these two sentences is surprising. Parfit is one of the people who has been most influential in arguing that, for Act Consequentialists, which moral beliefs *are true* is one question and which moral beliefs people *ought to have* is a different question (*Reasons and Persons*, pp. 40–41; *On What Matters*, vol. 3, pp. 415–16). Nevertheless, near the end of *On What Matters*, Parfit affirms that everyone ought to have the *same moral beliefs*, and he connects that idea with the idea that moral *truths* are universal.

What exactly is this connection between moral truth and sameness of belief? Parfit’s first remark, ‘everyone ought to have the same moral beliefs’, is about moral *belief*. His second remark, ‘moral truths are not true only for certain people’, is about moral *truth*. Presumably, the remark about universal moral truth is meant as justification for the remark about everyone’s sharing the same beliefs. On that presumption, I offer the following reconstruction of the reasoning behind the idea that everyone ought to have the same moral beliefs because moral truths are not true only for certain people:

- (a) Moral truths are not true only for certain people.
- (b) So moral truths are true for everyone.
- (c) If moral truths are true for everyone and if moral beliefs of some people conflict with those of other people, then at least some people’s moral beliefs must be false.
- (d) So, if moral truths are true for everyone, then at least some people’s moral beliefs must be false unless either everyone’s moral beliefs are the same or different people’s moral beliefs are different but do not conflict.
- (e) If moral truths are true for everyone, the only way for different people’s moral beliefs to be different and yet not conflict is for the different sets of moral beliefs to be incomplete. (For example, you have moral beliefs about A, B, C, D, E, F, G, but not

about H, I, or J, and I have moral beliefs about A, B, C, D, H, I, J, but not about E, F, and G.)

- (f) Although everyone cannot reasonably be required to have a complete set of true moral beliefs, everyone ought to have a *complete* set of *true* moral beliefs *about what the correct moral rules* are (and so about what properties of acts are morally positive or morally negative).<sup>20</sup>
- (g) If everyone has a *complete* set of *true* moral beliefs about what the correct moral rules are, then everyone has the *same* moral beliefs about what the correct moral rules are.
- (h) So everyone ought to have the *same* moral beliefs about what the correct moral rules are.

This argument is one I am suggesting might have somehow been in the background of Parfit's 'I am assuming everyone ought to have the same moral beliefs. Moral truths are not true only for certain people.'

Let me explain why premise (f) focuses on 'moral beliefs about what the correct moral rules are'. As (f) begins by acknowledging, any finite being cannot reasonably be required to have a complete set of true moral beliefs. A complete set of moral beliefs would have to include not only beliefs about the correct moral rules but also beliefs about all the more fundamental principles that singly or jointly ground the correct moral rules, or alternatively the belief that there is nothing more fundamental than the correct moral rules (the belief that is a large part of what I'm calling Rossian pluralist deontology). Parfit would have difficulty relying on a premise that includes the idea that everyone should have *true* moral beliefs about fundamental principles. That idea is in tension with Parfit's apparent acceptance that different people can accept different fundamental moral principles as long as there is convergence at the level of moral rules. Kantian Contractualists have one fundamental moral principle. Rule Consequentialists have a different one. Scanlonian Contractualists have yet another. Likewise for Virtue Ethicists. Rossian Pluralists deny there is any such single foundational principle. Parfit does not seem especially bothered by disagreement at the fundamental level.

That was an *ad hominem* argument against Parfit's having as a premise that everyone ought to have a complete set of true moral beliefs. Now I want to offer a second objection to the idea that everyone should have a complete set of true moral beliefs. This second objection should persuade everyone, not just Parfit. The objection is:

---

<sup>20</sup> Here I have been helped by a suggestion from Todd Karhu.

(P1) If everyone ought to have a *complete* set of *true* moral beliefs, then everyone ought to have a complete set of true moral beliefs about which actions are all-things-considered right or wrong in all particular cases.

(P2) It is not true that everyone ought to have a complete set of true moral beliefs about which actions are all-things-considered right or wrong in all particular cases.

So, it is not true that everyone ought to have a *complete* set of *true* moral beliefs.

This argument is logically valid. (P1) must be true, since, unless one has true beliefs about which specific acts are all-things-considered morally right in all particular cases, one does not have a *fully complete* set of true moral beliefs.

(P2) is also difficult not to accept. A complete set of true moral beliefs must include not only true moral beliefs about what is all-things-considered right in simple and familiar particular cases, but also true moral beliefs about what is all-things-considered right in complicated and unfamiliar cases. Complexity can arise from, among other things, complicated causal connections, conceptual subtleties, and conflicts between moral considerations neither of which obviously trumps the other. Cases might be unfamiliar because they occurred in the distant past or will occur in the changed circumstances of the future, or because they occur in possible worlds that haven't yet been considered. There is effectively no end of such actual or possible situations. Only an omniscient being could have a complete set of true moral beliefs about which actions are all-things-considered right or wrong in all particular cases. Since it is not true that everyone is omniscient, it cannot be true that everyone ought to have a complete set of true moral beliefs about which actions are all-things-considered right or wrong in all particular cases.

In short, because of the *variety* of possible situations, the *complexities* that can be morally pivotal, and the difficulty of judging which moral reasons outweigh other conflicting reasons in many cases, it is well beyond human capacity to have a complete set of true moral beliefs about what is all-things-considered right or wrong in all particular cases. So it is well beyond human capacity to have a complete set of true moral beliefs. Since this is something beyond human capacity, requiring people to achieve it would be preposterous.

Remember that Parfit's view that everyone ought to have the *same* moral beliefs is relevant to his rejection of the possibility of a mixed population in which some people have one set of moral beliefs (e.g. Act Consequentialist ones) and other people have a different set of moral beliefs (e.g. the beliefs that make up Common Sense Morality or Rule

Consequentialism). A world with such a mix might be one with greater aggregate welfare and other important goods. But, by the end of vol. 3 of *On What Matters*, Parfit dismisses such a world. He continues, ‘If everyone ought to have the same moral beliefs, we should ask whether things would on the whole go better if most of us were pure Act Consequentialists, or we accepted and tried to follow certain other moral rules.’ (p. 420) His answer is that things would on the whole go better if we accept the rules of some improved version of Common Sense Morality. And these are also the rules that Rule Consequentialism underwrites (pp. 421, 433).

### **Section 5: Conclusion**

In *Reasons and Persons*, Parfit appeared to think Act Consequentialism the best ethical theory. In vol. 1 of *On What Matters*, in contrast, Parfit argued for a Triple Theory combining Kantian Contractualism, Rule Consequentialism, and Scanlonian Contractualism. I have argued here that the reflective equilibrium methodology employed in assessing moral theories militates *not merely* in favour of a Triple Theory of Kantian Contractualism, Rule Consequentialism, and Scanlonian Contractualism, but in favour of a *Quintuple* Theory comprised of the best versions of Kantian Contractualism, Rule Consequentialism, Scanlonian Contractualism, Common Sense Morality, and Virtue Ethics.

In vol. 3 of *On What Matters*, Parfit devotes considerable space to exploring the capacity of what he called Act-involving Act Consequentialism to accord with our intuitions about the wrongness of some acts that would maximize utility. However, I have here marshalled some arguments against Act-involving Act Consequentialism.

One of Parfit’s arguments for Rule Consequentialism is a reflective equilibrium argument. The other is that, with respect to a very important class of cases, Rule Consequentialism’s explanation of wrongness in terms of ‘the combined effects of what we and others together do’ is superior to Act Consequentialism’s explanation in terms of the effects of an individual’s act. I emphatically endorse these arguments, but I have not here tried to buttress them.

A third argument Parfit offers for Rule Consequentialism is surprising. This argument compares a world in which everyone accepts Act Consequentialism with a world in which everyone accepts an improved version of Common Sense Morality or Rule Consequentialism. Why not add to the comparison a world in which some people accept Act Consequentialism and other people accept Common Sense Morality or Rule Consequentialism? Parfit’s answer

is that everyone ought to have the same moral beliefs since moral truths are the same for everyone. I have tried to reconstruct what might be the thinking behind that answer.<sup>21</sup>

## Bibliography

- Bales, R.E., ‘Act-utilitarianism: Account of Right-making Characteristics or Decision-making Procedure?’, *American Philosophical Quarterly* 8/3 (1971): 257–65.
- Baumann, M., ‘Parfit, Convergence, and Underdetermination’, *Journal of Ethics and Social Philosophy* 13/3 (2018): 191–221.
- Broad, C.D., *Ethics*, C. Lewy (ed.), (Nijhoff Publishers, 1985).
- Copp, D., ‘The Rule Worship and Idealization Objections Revisited and Resisted’, in M. Timmons (ed.), *Oxford Studies in Normative Ethics* 10 (2020).
- Cullity, G., *The Moral Demands of Affluence* (Oxford University Press, 2004).
- Frick, J., ‘Contractualism and Social Risk’, *Philosophy and Public Affairs* 43/3 (2015): 175–223.
- Harrod, R.F., ‘Utilitarianism Revised’, *Mind* 45/178 (1936): 137–56.
- Hooker, B., ‘Rule-consequentialism, Incoherence, Fairness’, *Proceedings of the Aristotelian Society* 95/1 (1995): 19–35.
- ‘Ross-style Pluralism versus Rule-consequentialism’, *Mind* 105/420 (1996): 531–52.
- ‘Reflective Equilibrium and Rule Consequentialism’, in B. Hooker, E. Mason, and D. Miller, eds., *Morality, Rules, and Consequences* (Edinburgh University Press, 2000): 221–238.
- *Ideal Code, Real World: A Rule-consequentialist Theory of Morality* (Oxford University Press, 2000).
- ‘Rule-consequentialism and Internal Consistency: A Reply to Card’, *Utilitas* 19/4 (2007): 514–19.
- ‘Must Kantian Contractualism and Rule-consequentialism Converge?’, in Mark Timmons, (ed.), *Oxford Studies in Normative Ethics* 4 (2014): 34–52.
- Mulgan, T., *Future People: A Moderate Consequentialist Account of Our Obligations to Future Generations* (Oxford University Press, 2003).
- Murphy, L., *Moral Demands in Nonideal Theory* (Oxford University Press, 2000).
- Parfit, D., *Reasons and Persons*, (Oxford University Press, 1984).
- *On What Matters*, volumes 1 & 2 (Oxford University Press, 2011).
- *On What Matters*, volume 3 (Oxford University Press, 2017).
- Portmore, D., *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford University Press, 2011).
- Ross, W.D., *The Right and the Good* (Oxford University Press, 1930).
- Rüger, K., ‘On Ex Ante Contractualism’, *Journal of Ethics and Social Philosophy* 13/3 (2018): 240–258.
- Scanlon, T.M., *What We Owe to Each Other* (Harvard University Press, 1998).
- Scheffler, S., *The Rejection of Consequentialism* (Oxford University Press, 1982).

---

<sup>21</sup> This paper has been presented at conferences at York, Reading, and Karlsruhe. The paper has been revised to deal with insightful remarks from Marco Bernardini, James Brown, Jamie Draper, Johan Gustafsson, Matthias Hoesch, Chris Jay, Todd Karhu, Kacper Kowalczyk, Katarzyna de Lazari-Radek, Giuliana Manca, Tim Mulgan, Taylor Paytas, Adam Pearce, Chris Ryder, Christian Seidel, José de Sousa E. Brito, Rob Truman, Elizabeth Ventham, Tatjana Visak, Hayden Wilkinson, and Michael Zimmerman. I am grateful to all these people. For extremely helpful written comments, I thank Karhu, Rob Jubb, Roger Crisp, Aart van Gils, Charlotte Newey, Jeff McMahan, and Peter Vallentyne.

Sidgwick, H., *Methods of Ethics*, 7<sup>th</sup> edition (Macmillan, 1907).

Urmson, J.O., 'On the Interpretation of the Moral Philosophy of J. S. Mill', *Philosophical Quarterly* 3/10 (1953): 33–9.

Wolf, S., 'Two Concepts of Rule Utilitarianism', in M. Timmons (ed.), *Oxford Studies in Normative Ethics* 6 (2016): 123-144.