REVIEW OF *ON WHAT MATTERS, VOLUME 1*

1. Summary

Parfit wants to show that three major ethical theories—Kantianism, contractualism, and consequentialism—actually converge toward a single unified theory. To do this, he looks at some prominent versions of those theories, points out weaknesses in them, and argues that those weaknesses are best addressed by revising the theories in particular ways. Parfit thinks these revisions make it so that the theories often or even always agree with each other, and can therefore be combined into what he calls the "**Triple Theory**":

> An act is wrong just when such acts are disallowed by some principle that is optimific, uniquely universally willable, and not reasonably rejectable.[1]

I think there are really two core premises which drive Parfit's revisions to the original theories and which, if accepted, make the convergence plausible:

- Morality must reflect a certain kind of **impartiality**.
- **Objective reasons** exist and morality must be informed by them.

2. Impartiality

Parfit's view leaves plenty of room for moral principles that allow people to prioritize their selfish interests, or the interests of their loved ones, over the interests of strangers. The principles themselves, however, should not be chosen in a way that stacks the deck in favor of particular people.

Thus Parfit rejects, for example, versions of contractualism that ask merely what kind of contract *real* people *would actually* make if they all sat down and tried to agree on a set of moral rules. In such a process, more privileged people would have more bargaining power. Parfit summarizes a better version of contractualism, devised by Scanlon, as follows: "Everyone ought to follow the principles that no one could reasonably reject."[2] The word *reasonable* is meant in a "partly moral sense"[3] which seems to bake a degree of impartiality into the theory: "We are unreasonable in this sense if we give too little weight to other people's well-being or moral claims."[4] (I wrote a detailed walkthrough of Scanlon's book *What We Owe to Each Other* here.)

Similarly, Parfit rejects forms of Kantianism that ask merely what *you* could will to be a universal law. Many people can will things to be law which place unfair burdens on others, simply because they themselves are unlikely to suffer the consequences. Parfit develops what he calls Kantian Contractualism: "Everyone ought to follow the principles that *everyone* could rationally will to be universal laws."[5] (If you didn't catch the difference, consider: I could wish that everyone followed the rule *give Jacob all your money every time you see him*. But not everyone could wish that everyone followed that rule.)

3. Reasons

On a *subjectivist* view, "our reasons for acting are all provided by, or depend upon, certain facts about what would fulfil or achieve our present desires or aims."[6] In contrast, Parfit accepts an *objectivist* view (not to be confused with Ayn Rand's so-called "Objectivism"), in which "there are certain facts that give us reasons both to have certain desires and aims, and to do whatever might achieve these aims."[7]

The notion of objective reasons does a lot of work in this book. It allows Parfit to call things "universally willable" or "not reasonably rejectable" even when the individuals involved *don't* will them and *would* reject them. For some conflicts, there is no solution that everyone involved *would actually* accept, nor even a solution that is compatible with everyone's most fundamental desires/goals/interests. But if we instead look for a solution everyone "could rationally" accept—where *rationally* is defined with respect to objective reasons—we may be able to find one by appealing to claims about what people *should* be willing to accept. Consider Parfit's defense of a "Consent Principle", where he discusses the following example:

> …two people, *White* and *Grey*, are trapped in slowly collapsing wreckage. I am a rescuer, who could prevent this wreckage from either killing White or destroying Grey's leg.[8]

Parfit thinks that "[i]f these are the only morally relevant facts, it is clear that I ought to save White's life."[9] And he thinks this conclusion can be justified in terms of "sufficient reasons to consent"[10].

> Grey could rationally choose that I save her leg, since this choice would be much better for her. But she would not be rationally required to make this choice. Grey could rationally choose instead that I save White's life. Grey could rationally regard White's well-being as mattering about as much as hers, and White's loss in dying would be much greater than Grey's loss in losing her leg.
>
> White, in contrast, could not rationally choose that I save Grey's leg. We could often rationally choose to benefit some stranger, I believe, even if our choice would make us lose a somewhat greater benefit. But there is too great a difference between the possible benefits to White and Grey. White would not have sufficient reasons to give up her life so that I could save Grey's leg.[11]

Imagine a Venn diagram showing which option(s) each person could rationally endorse; to find a morally acceptable option, you look in the region of overlap.

outcomes Grey could rationally choose
outcomes White could rationally choose
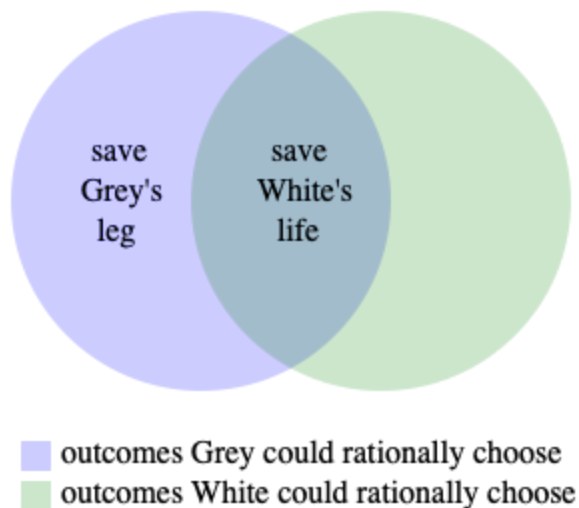
(Of course, Parfit also recognizes that "[i]t is often morally important whether people *actually* consent to being treated in some way, or whether, if they had the opportunity, these people would in fact consent."[12])

4. Motivation

By showing a convergence among different moral theories, Parfit ultimately hoped to neutralize one of the arguments against morality in general:

> Of our reasons for doubting that there are moral truths, one of the strongest is provided by some kinds of **moral disagreement**. Most moral disagreements do not count strongly against the belief that there are moral truths, since these disagreements depend on different people's having conflicting empirical or religious beliefs, or on their having conflicting interests, or on their using different concepts, or these disagreements are about borderline cases, or they depend on the false assumption that all questions must have answers, or precise answers. But some disagreements are not of these kinds. These disagreements are deepest when we are considering, not the wrongness of particular acts, but the nature of morality and moral reasoning, and what is implied by different views about these questions. If we and others hold conflicting views, and we have no reason to believe that *we* are the people who are more likely to be right, that should at least make us doubt our view. It may also give us reasons to doubt that any of us could be right.[13]

He was, at least according to Edmonds's biography of him (review), fanatically obsessed with this issue. At stake for Parfit was not just a particular conception of morality, but *whether anything matters at all*:

> …Parfit came to believe that dissent about ethics—especially dissent between leading philosophers—was evidence for its relativism. And he thought that relativism essentially collapsed into nihilism. If your moral truth conflicted with, but was no less valid than, my moral truth, this would show that, **ultimately, nothing mattered**.[14]

I share this worry. There are, perhaps, two levels to it. The prospect that all morality might be arbitrary—that there is no meaningful sense in which, for example, someone who likes to torture children is worse than someone who likes to nurture them—would be horrifying enough. But what really motivates that sort of relativism seems, generally, to be subjectivism about reasons, which can undermine our values in an even more radical way. Subjectivism would (in my opinion) imply not only that our goals and desires are fundamentally arbitrary, but also that the whole project of trying to accomplish goals and fulfill desires is arbitrary. Seeking one's own success and happiness would be no more worthwhile than seeking failure and misery, because there would be no facts about what's worthwhile at all—just facts about what we happen to be programmed to do.

But Parfit's anxiety over the issue doesn't really seem rational to me. For one thing, the strongest argument against the conclusion that "nothing matter[s]" is just that some things obviously *do* matter. It's self-evident to anyone who's experienced intense suffering that it matters how much more of it they have to suffer. We can, I think, perceive this fact much more vividly and directly than we can perceive the validity of any purported argument against it.

But also, most people who accept subjectivism and relativism are *not* nihilists. So Parfit (and I) disagree with them on two issues: which theory is true, and what the implications of each theory are. In disagreements like this, it's worth keeping in mind that if we were persuaded to change our minds on one issue, we would likely acquire a new perspective on the other issue too. (To give a related example: evangelical Christians are often convinced that God is necessary to justify morality. But when they deconvert, they don't become amoral; instead, they become convinced that morality has a different foundation, and often become convinced that postulating a God wouldn't even *help* with grounding morality anyway.)

5. Evaluation

Summarizing the reception of this book, Edmonds wrote:

> …the general tenor of the reviews was that Parfit's project resembled a **vast baroque cathedral** that evoked a sense of awe less for its beauty than for its sheer construction. 'It stands as a grand and dedicated attempt to elaborate a fundamentally misguided

> perspective,' declared *The New Republic*. Several of the reviews mentioned the daunting length of the volumes; one reviewer went to the trouble of putting Volumes 1 and 2 on the scales: they weighed in at '4.8 pounds' (2.18 kilos).[15]

That was pretty much my impression when I first read the book a decade ago: too long and mostly pointless.

On rereading, I noticed that Parfit uses whitespace as generously as an undergraduate trying to meet a page-count requirement, so volume 1 isn't really as long as it looks.

More importantly, I have more appreciation for the problem of disagreement that gave Parfit so much anxiety. When I was younger and more arrogant, I was perhaps more prone to dismiss people who disagreed with me, on the assumption that they'd given the issue less careful thought or simply weren't as intelligent.

Does Parfit successfully defang the problem of moral disagreement? Doubtful. But I think he shows that quite a lot of convergence is plausible if you accept the existence of objective reasons. This suggests the deepest point of disagreement in ethics may really be the subjectivist vs objectivist debate.

(I think the whole *point* of Korsgaard's version of Kantianism, for example, was to avoid relying on any notion of objective reasons.)

If we accept objectivism, the next big question is just what range of objective reasons actually exist. It's easiest to make the case that we have reasons to avoid suffering and pursue happiness; but if you stop there, I think you just get utilitarianism. Parfit's arguments seem to depend on humans having a more open-ended capacity for recognizing objective reasons.

6. Aside: Four senses of "wrong"

Parfit is great at drawing subtle distinctions that allow us to analyze a situation more clearly. For example, I really like how chapter 7 ("Moral Concepts") distinguishes several senses in which an action might be called (morally) "wrong":

> It is often assumed that the word 'wrong' has only one moral sense. This assumption is most plausible when we are considering the acts of people who know all of the morally relevant facts. We can start by supposing that, when we think about such acts, we all use 'wrong' in the same sense, which we can call the *ordinary* sense. In many cases, however, we don't know all of the relevant facts, and we must act in ignorance, or with false beliefs. When we think about such cases, we can use 'wrong' in several partly different senses. Some of these senses we can define by using the ordinary sense. Some act of ours would be
>
>   *wrong* in the *fact-relative* sense just when this act would be wrong in the ordinary sense if we knew all of the morally relevant facts,
>
>   *wrong* in the *belief-relative* sense just when this act would be wrong in the ordinary sense if our beliefs about these facts were true,
>
> and
>
>   *wrong* in the *evidence-relative* sense just when this act would be wrong in the ordinary sense if we believed what the available evidence gives us decisive reasons to believe, and these beliefs were true.[16]
>
> … [a few pages later]

> *wrong* in the *moral-belief-relative* sense just when the agent believes this act to be wrong in the ordinary sense.[17]

1. Derek Parfit, *On What Matters*, The Berkeley Tanner Lectures (Oxford ; New York: Oxford University Press, 2011), 413.

2. Ibid., 360.

3. Ibid.

4. Ibid.

5. Ibid., 407, emphasis added.

6. Ibid., 45.

7. Ibid.

8. Ibid., 185.

9. Ibid.

10. Ibid., 184.

11. Ibid., 186.

12. Ibid., 191.

13. Ibid., 418–19, emphasis added.

14. David Edmonds, *Parfit: A Philosopher and His Mission to Save Morality* (Princeton: Princeton University Press, 2023), 280, emphasis added.

15. Ibid., 293, emphasis added.

16. Parfit, *On What Matters*, 150–51.

17. Ibid., 158.