

# Arguments for Utilitarianism

## Table of Contents

1. Introduction: Moral Methodology & Reflective Equilibrium
2. Arguments for Utilitarianism
  - What Fundamentally Matters
  - The Golden Rule, the Veil of Ignorance, and the Ideal Observer
  - Expanding the Moral Circle
3. Arguments for Consequentialism
  - Agency as a Force for Good
  - The Poverty of the Alternatives
4. Conclusion
5. Resources and Further Reading

## Introduction: Moral Methodology & Reflective Equilibrium

You cannot *prove* a moral theory. Whatever arguments you come up with, it is always possible for someone else to reject your premises—if they are willing to accept the costs of doing so. Different theories offer different advantages. This chapter will set out some of the major considerations that plausibly count in favor of utilitarianism. A complete view also needs to consider the costs of utilitarianism (or the advantages of its competitors), which is addressed in the chapter [Objections to Utilitarianism](#). You can then reach an all-things-considered judgment as to which moral theory strikes you as overall best or most plausible.

To this end, moral philosophers typically use the methodology of *reflective equilibrium*.<sup>1</sup> This involves balancing two broad kinds of evidence as applied to moral theories:

1. Intuitions about specific cases (thought experiments).
2. General theoretical considerations, including the plausibility of the theory's *principles* or systematic claims about what matters.

General principles can be challenged by coming up with putative *counterexamples*, or cases in which they give an intuitively incorrect verdict. In response to such putative counterexamples, we must weigh the force of the case-based intuition against the inherent plausibility of the principle being challenged. This could lead you to *either* revise the principle to accommodate your intuitions

about cases *or* to reconsider your verdict about the specific case, if you judge the general principle to be better supported (especially if you are able to “explain away” the opposing intuition as resting on some implicit mistake or confusion).

As we will see, the arguments in favor of utilitarianism rest overwhelmingly on general theoretical considerations. [Challenges to the view](#) can take either form, but many of the most pressing objections involve thought experiments in which utilitarianism is held to yield counterintuitive verdicts.

There is no neutral, non-question-begging answer to how one ought to resolve such conflicts.<sup>2</sup> It takes judgment, and different people may be disposed to react in different ways depending on their philosophical temperament. As a general rule, those of a temperament that favors *systematic theorizing* are more likely to be drawn to utilitarianism ([and related views](#)), whereas those who hew close to common sense intuitions are less likely to be swayed by its theoretical virtues. Considering the arguments below may thus do more than just illuminate utilitarianism; it may also help you to discern your own philosophical temperament!

While our presentation focuses on utilitarianism, it is worth noting that many of the arguments below could also be taken to support [other forms of welfarist consequentialism](#) (just as many of the [objections to utilitarianism](#) also apply to these related views). This chapter explores arguments for utilitarianism and closely related views over non-consequentialist approaches to ethics.

## Arguments for Utilitarianism

### What Fundamentally Matters

Moral theories serve to specify *what fundamentally matters*, and utilitarianism seems to offer a particularly compelling answer to this question.

Almost anyone would agree with utilitarianism that suffering is bad, and happiness is good. What could be more obvious? If anything matters morally, human well-being surely does. And it would be arbitrary to limit moral concern to our own species, so we should instead conclude that [well-being](#) generally is what matters. That is, we ought to want the lives of sentient beings to go as well as possible (whether that ultimately comes down to maximizing [happiness](#), [desire satisfaction](#), or [other welfare goods](#)).

Could anything else be *more* important? Such a suggestion can seem puzzling. Consider: it is (usually) wrong to steal.<sup>3</sup> But that is plausibly because stealing tends to be *harmful*, reducing people’s well-being.<sup>4</sup> By contrast, more people are open to redistributive taxation, if it allows governments to provide benefits that reliably raise the overall level of well-being in society. So it is not that individuals just have a natural right to not be interfered with no matter what. When judging institutional arrangements (such as property and tax law), we recognize that what matters

is coming up with arrangements that tend to secure *overall good results*, and that the most important factor in what makes a result *good* is that it *promotes well-being*.<sup>5</sup>

Such reasoning may justify viewing utilitarianism as the default starting point for moral theorizing. If someone wants to claim that there is some other moral consideration that can override *overall well-being* (trumping the importance of saving lives, reducing suffering, and promoting flourishing), they face the challenge of explaining *how* that could possibly be so. Many common moral rules (like those that prohibit theft, lying, or breaking promises), while not explicitly utilitarian in content, nonetheless have a clear utilitarian rationale. If they did not generally promote well-being—but instead actively harmed people—it is hard to see what reason we would have to still want people to follow them. To follow *harmful* moral rules would seem like a kind of “rule worship”, and not truly ethical at all.<sup>6</sup>

Similar judgments apply to hypothetical cases in which you somehow know for sure that a typically reliable rule is, in this particular instance, counterproductive. In the extreme case, we all recognize that you ought to lie or break a promise if lives are on the line. In practice, of course, the best way to achieve good results over the long run is to [respect commonsense moral rules and virtues](#) while seeking opportunities to help others. (It is important not to mistake the hypothetical verdicts utilitarianism offers in stylized thought experiments with [the practical guidance it offers in real life](#).) The key point is that utilitarianism offers a seemingly unbeatable answer to the question of *what fundamentally matters*: protecting and promoting the interests of all sentient beings to make the world as good as it can be.

## The Golden Rule, the Veil of Ignorance, and the Ideal Observer

Humans are masters of self-deception and motivated reasoning. If something benefits us personally, it is all too easy to convince ourselves that it must be okay. To correct for such self-serving biases, it can be helpful to consider [moral heuristics](#) like the Golden Rule: “*do unto others as you would have them do unto you*.”

Many ethical traditions endorse some form of the Golden Rule. Of course, it needs to be interpreted sensibly: we would not want a masochist to go around whipping people who are not so into that as he is. The idea underlying the Golden Rule is arguably the *equal consideration of interests*: that we “give equal weight in our moral deliberations to the like interests of all those affected by our actions,”<sup>7</sup> just as we would want others to do with our interests.

One vivid way to do this is to imagine yourself in the position of *each* affected person, one after the other, living each of their lives (with *their* tastes and preferences) in sequence. If you were, in effect, *everybody*, what would it be rational for you to choose? [R. M. Hare](#) argued that the answer was *preference utilitarianism*: to best satisfy everyone’s overall preferences, counting everyone equally.

Of course, if you accept a different [theory of well-being](#),<sup>8</sup> then you might think it worth overriding misguided or irrational preferences for the sake of greater happiness or other forms of flourishing. But it remains the case that some form of utilitarianism seems likely to follow from such a procedure of putting yourself into every person's shoes.

Similar results may be obtained by instead imagining that you are looking down on the world from behind a “veil of ignorance” that reveals the facts about each person's circumstances in society, while hiding from you the knowledge of *which of these individuals you are*.<sup>9</sup> Imagine you were trying to decide on the best structure for society from behind this veil of ignorance. If you view yourself as equally likely to end up being anyone in the world, it would seem prudent to maximize overall well-being, just as utilitarianism prescribes.<sup>10</sup>

Finally, impartiality might be secured by simply asking what a suitably disinterested but benevolent observer would want. For example, whether or not one believes that any God exists, one might consider the hypothetical question of what an ideal (all knowing and all loving) observer *would* want—as it is plausible that the answer to this question would also settle what we *ought* to do. But then, it seems hard to deny that a benevolent impartial observer would simply want whatever would be best overall, that is counting everyone's interests equally. And that in turn means that it is hard to deny utilitarianism.

## Expanding the Moral Circle

When we look back on past moral atrocities—like slavery or denying women equal rights—we recognize that they were often sanctioned by the dominant societal norms at the time. The perpetrators of these atrocities were grievously wrong to exclude their victims from their “circle” of moral concern.<sup>11</sup> That is, they were wrong to be indifferent towards (or even delight in) their victims' suffering. But such exclusion seemed normal to people at the time. So we should question whether we might likewise be blindly accepting of some practices that future generations will see as evil but that seem “normal” to us.<sup>12</sup> The best protection against making such an error ourselves would be to deliberately expand our moral concern outward, to include *all* sentient beings—anyone who can suffer—and so recognize that we have strong moral reasons to reduce suffering and promote well-being wherever we can, no matter *who* it is that is experiencing it.

While this conclusion is not yet all the way to full-blown utilitarianism, since it is compatible with, for example, holding that there are side-constraints limiting one's pursuit of the good, it is likely sufficient to secure agreement with the most important [practical implications of utilitarianism](#) (stemming from [cosmopolitanism](#), [anti-speciesism](#), and [longtermism](#)).

## Arguments for Consequentialism

Utilitarianism is the paradigmatic form of consequentialism. So one way to support utilitarianism (alongside other consequentialist views) is to argue in favor of consequentialism more broadly, and

against the non-consequentialist alternatives. The remainder of this chapter explores such arguments.

## Agency as a Force for Good

If some form of consequentialism is true, then we can expect (successful)<sup>13</sup> moral agents to have positive instrumental value: their presence in a situation will tend to make things go better. On non-consequentialist views, by contrast, there is no such connection between doing the right thing and making things better. Sometimes agents might even be obliged to make things *worse* (because the better *outcome* is only obtainable through making a morally worse *choice*, such as wrongly killing an innocent person). Insofar as we are inclined to think that moral agents should be a force for good in the world, that gives us reason to favor consequentialist views.

It can be difficult to resist the general consequentialist principle that more desirable outcomes are more worth bringing about. Put simply: if you have a choice between bringing about a better outcome or a worse one, surely it would be morally better to choose the better outcome?<sup>14</sup>

This argument rests on a conception of action as *teleological* or goal-directed in nature. The purpose of action, on this view, is to *bring about* desired outcomes. Of course, we are not always so deliberate. One may spontaneously sing a song without intending to bring about any *further* effect. But it remains plausible that this is a good idea only if the outcome which includes one's singing is better (perhaps because it is more joyful) than the available alternatives. After all, if remaining silent would somehow save another's life, that would surely change things. So it is hard to deny that actions are justified to the extent that they bring about things that we ought to want—that is, good outcomes.

On this view, if it would be good for something to happen, then it would be good, all else equal, to *choose* this outcome—to *make* it happen. Consider a “lifeboat” scenario where five people are drowning in one direction, and one person is drowning in the opposite direction, with not enough time to save everyone. We all agree it would be better if more lives are saved—for instance, if a lifeboat, floating in the water, were to drift towards the five rather than the one. If we now add a captain to direct the lifeboat, some non-consequentialists would instead insist that fairness requires the captain to flip a coin to decide which group to save. But that violates the principle that better outcomes are more worth choosing.

It would seem strange to think that giving a fully informed and morally ideal agent control over a situation could make the outcome worse than if they had not been there at all. But that can sometimes happen on non-consequentialist accounts, which sometimes require agents to choose (or needlessly risk) worse outcomes.

To see this, suppose that (i) the lifeboat would naturally drift towards the five, if no-one was steering it, and yet (ii) because the captain is steering the boat, fairness requires him to flip a coin

to decide which group to save. If both claims were true, then it would be a bad thing for the world to have the captain steering the lifeboat, even when he is fully informed and morally ideal. For that would needlessly risk the worse outcome (depending on the flip of the coin). We would be better off just letting the lifeboat drift naturally towards the five, ensuring the best outcome. But if that is better, then it seems that a morally ideal agent ought to prefer that, too. This casts doubt on the non-consequentialist claim (ii), above, that fairness requires the captain to flip a coin and risk a worse outcome. Morally ideal agents should bring about the outcome that they ought to prefer—namely, the outcome that is overall best.

Non-consequentialist views imply that it could be “wrong” to make the world a better place, and “right” to make it worse. But that seems odd. Worse, it implies that it could be *undesirable* to have more moral agents making correct decisions. Only consequentialist views, such as utilitarianism, ensure that it is always morally right to bring about better outcomes, and hence that correct moral decision-making is a force for good in the world.

## The Poverty of the Alternatives

We have seen that there is a strong presumptive case in favor of utilitarianism. If no competing view can be shown to be superior, then utilitarianism has a strong claim to be the “default” moral theory that we should fall back on. It turns out that one of the strongest considerations in favor of utilitarianism (and related consequentialist views) is the deficiencies of the alternatives.

Deontological (or rule-based) theories, in particular, seem to rest on questionable foundations.<sup>15</sup>

Deontological theories are explicitly *non-consequentialist*: instead of morally assessing actions by evaluating their consequences, these theories tend to take certain types of action (such as killing an innocent person) to be *intrinsically* wrong.<sup>16</sup> There are reasons to be dubious of this approach to ethics, however.

## The Paradox of Deontology

Deontologists hold that there is a *constraint* against killing: that it is wrong to kill an innocent person even if this would save five *other* innocent people from being killed. This verdict can seem puzzling on its face.<sup>17</sup> After all, given how terrible killing is, should we not want there to be *less* of it? As Scheffler (1985) argued, rational choice in general tends to be goal-directed, a conception which fits poorly with deontic constraints.<sup>18</sup> A deontologist might claim that their goal is simply to avoid violating moral constraints *themselves*, which they can best achieve by not killing anyone, even if that results in more individuals being killed. While this explanation can render deontological verdicts coherent, it does so at the cost of making them seem awfully narcissistic, as though the deontologist’s central concern was just to maintain their own moral purity or “clean hands”.

Deontologists might push back against this characterization by instead insisting that moral action need not be goal-directed at all.<sup>19</sup> Rather than only seeking to promote value (or minimize harm), they claim that moral agents may sometimes be called upon to *respect* another's value (by not harming them, even as a means to preventing greater harm to others), which would seem an appropriately outwardly-directed, non-narcissistic motivation.

Scheffler's challenge remains that such a proposal makes moral norms puzzlingly divergent from other kinds of practical norms. If morality sometimes calls for respecting value rather than promoting it, why is the same not true of prudence? (Given that pain is bad for you, for example, it would not seem prudent to refuse a painful operation now if the refusal commits you to five comparably painful operations in future.) Deontologists may offer various answers to this question, but insofar as we are inclined to think, pre-theoretically, that ethics ought to be continuous with other forms of rational choice, that gives us some reason to prefer consequentialist accounts.

### The Hope Objection

We earlier suggested that impartial observers should want and hope for the best outcome. Non-consequentialists claim that nonetheless it is sometimes wrong to bring about the best outcome. Putting the two claims together yields the striking result that you should sometimes hope that others act wrongly (even when they are fully informed).

Suppose it would be wrong for some stranger—call him Jack—to kill one innocent person to prevent five other (morally comparable) killings. Non-consequentialists may claim that Jack has a special responsibility to ensure that *he* does not kill anyone, even if this results in more killings by others. But *you* are not Jack. From your perspective as an impartial observer, Jack's killing one innocent person is no more or less intrinsically bad than any of the five other killings that would thereby be prevented. You have most reason to hope that there is only one killing rather than five. So you have reason to hope that Jack acts “wrongly” (killing one to save five). But that seems odd.

More than merely being odd, this might even be taken to undermine the claim that deontic constraints *matter*, or are genuinely *important* to abide by. After all, to be important just is to be worth caring about. For example, we should care if others are harmed, which validates the claim that others' interests are morally important. But if we should not care more about Jack's abiding by the moral constraint against killing than we should about his saving five lives, that would seem to suggest that the constraint against killing is *not* in fact more morally important than saving five lives.

Finally, since our moral obligations ought to track what is genuinely morally important, if deontic constraints are not in fact important then we cannot be obligated to abide by them.<sup>20</sup> We cannot be obliged to prioritize deontic constraints over others' lives, if we ought to care more about others' lives than about deontic constraints. So deontic constraints must not accurately describe our

obligations after all. Jack really ought to do whatever would do the most good overall, and so should we.

### Skepticism About the Doing vs. Allowing Distinction

You might wonder: if respect for others requires not harming them (even to help others more), why does it not equally require not *allowing* them to be harmed? Deontological moral theories place great weight on distinctions such as those between [doing and allowing harm](#), or killing and letting die, or intended versus merely foreseen harms. But *why* should these be treated so differently? If a victim ends up equally dead either way, whether they were killed or “merely” allowed to die would not seem to make much difference to them—surely what matters to them is just their death.

Indeed, it is far from clear that there *is* any robust distinction between “doing” and “allowing”. Sometimes you might “do” something by remaining perfectly still.<sup>21</sup> Also, when a doctor unplugs a terminal patient from life support machines, this is typically thought of as “letting die”; but if a mafioso, worried about an informant’s potentially incriminating testimony, snuck in to the hospital and unplugged the informant’s life support, we are more likely to judge it to constitute “killing”.<sup>22</sup> Bennett (1998) argues at length that there is no satisfactory, fully general distinction between doing and allowing—at least, none that would vindicate the moral significance that deontologists want to attribute to such a distinction.<sup>23</sup> If Bennett is right, then that might force us towards some form of consequentialism (such as utilitarianism) instead.

### Status Quo Bias

Opposition to utilitarian trade-offs—that is, benefiting some at a lesser cost to others—amounts to a kind of status quo bias, prioritizing the *preservation of privilege* over promoting well-being more generally.

Such conservatism might stem from the Just World fallacy: the mistake of assuming that the status quo is just, and that people naturally get what they deserve. Of course, reality offers no such guarantees of justice. What circumstances one is born into depends on sheer luck, including one’s endowment of physical and cognitive abilities which may pave the way for future success or failure. Thus, even later in life we never manage to fully wrest back control from the whimsies of fortune and, consequently, some people are vastly better off than others despite being no more deserving. In such cases, why should we not be willing to benefit one person at a lesser cost to privileged others? They have no special entitlement to the extra well-being that fortune has granted them.<sup>24</sup> Clearly, it is good for people to be well-off, and we certainly would not want to harm anyone unnecessarily.<sup>25</sup> However, if we can increase overall well-being by benefiting one person at the lesser cost to another, we should not refrain from doing so merely due to a prejudice in favor of the existing distribution.<sup>26</sup> It is easy to see why traditional elites would want to promote a “morality”



which favors their entrenched interests. It is less clear why others should go along with such a distorted view of what (and who) matters.

It can similarly be argued that there is no real distinction between imposing harms and withholding benefits. The only difference between the two cases concerns what we understand to be the status quo, which lacks moral significance. Suppose scenario A is better for someone than B. Then to shift from A to B would be a “harm”, while to prevent a shift from B to A would be to “withhold a benefit”. But this is merely a descriptive difference. If we deny that the historically given starting point provides a morally privileged baseline, then we must say that the cost in either case is the same, namely the difference in well-being between A and B. In principle, it should not matter where we start from.<sup>27</sup>

Now suppose that scenario B is vastly better for someone else than A is: perhaps it will save their life, at the cost of the first person’s arm. Nobody would think it okay to kill a person just to save another’s arm (that is, to shift from B to A). So if we are to avoid status quo bias, we must similarly judge that it would be wrong to *oppose* the shift from A to B—that is, we should not object to saving someone’s life at the cost of another’s arm.<sup>28</sup> We should not care especially about preserving the privilege of whoever stood to benefit by default; such conservatism is not truly fair or just. Instead, our goal should be to bring about whatever outcome would be best *overall*, counting everyone equally, just as utilitarianism prescribes.

### Evolutionary Debunking Arguments

Against these powerful theoretical objections, the main consideration that deontological theories have going for them is closer conformity with our intuitions about particular cases. But if these intuitions cannot be supported by independently plausible principles, that may undermine their force—or suggest that we should interpret these intuitions as good rules of thumb for practical guidance, rather than as indicating what *fundamentally* matters.

The force of deontological intuitions may also be undermined if it can be demonstrated that they result from an unreliable process. For example, evolutionary processes may have endowed us with an emotional bias favoring those who look, speak, and behave like ourselves; this, however, offers no justification for discriminating against those unlike ourselves. Evolution is a blind, amoral process whose only “goal” is the propagation of genes, not the promotion of well-being or moral rightness. Our moral intuitions require scrutiny, especially in scenarios very different from our evolutionary environment. If we identify a moral intuition as stemming from our evolutionary ancestry, we may decide not to give much weight to it in our moral reasoning—the practice of *evolutionary debunking*.<sup>29</sup>

Katarzyna de Lazari-Radek and Peter Singer argue that views permitting partiality are especially susceptible to evolutionary debunking, whereas [impartial](#) views like utilitarianism are more likely

to result from undistorted reasoning.<sup>30</sup> Joshua Greene offers a different psychological debunking argument. He argues that deontological judgments—for instance, in response to [trolley cases](#)—tend to stem from unreliable and inconsistent emotional responses, including our favoritism of identifiable over faceless victims and our aversion to harming someone up close rather than from afar. By contrast, utilitarian judgments involve the more deliberate application of widely respected moral principles.<sup>31</sup>

Such debunking arguments raise worries about whether they “prove too much”: after all, the foundational moral judgment that *pain is bad* would itself seem emotionally-laden and susceptible to evolutionary explanation—physically vulnerable creatures would have powerful evolutionary reasons to want to avoid pain *whether or not* it was objectively bad, after all!

However, debunking arguments may be most applicable in cases where we feel that a principled explanation for the truth of the judgment is lacking. We do not tend to feel any such lack regarding the badness of pain—that is surely an intrinsically plausible judgment if anything is. Some intuitions may be *over-determined*: explicable *both* by evolutionary causes *and* by their rational merits. In such a case, we need not take the evolutionary explanation to undermine the judgment, because the judgment *also* results from a reliable process (namely, rationality). By contrast, deontological principles and partiality are far less *self-evidently* justified, and so may be considered more vulnerable to debunking. Once we have an explanation for these psychological intuitions that can explain why we would have them even if they were rationally baseless, we may be more justified in concluding that they are indeed rationally baseless.

As such, debunking objections are unlikely to change the mind of one who is drawn to the target view (or regards it as independently justified and defensible). But they may help to confirm the doubts of those who already felt there were some grounds for scepticism regarding the intrinsic merits of the target view.

## Conclusion

Utilitarianism can be supported by several theoretical arguments, the strongest perhaps being its ability to capture *what fundamentally matters*. Its main competitors, by contrast, seem to rely on dubious distinctions—like “doing” vs. “allowing”—and built-in status quo bias. At least, that is how things are apt to look to one who is broadly sympathetic to a utilitarian approach. Given the flexibility inherent in reflective equilibrium, these arguments are unlikely to sway a committed opponent of the view. For those readers who find a utilitarian approach to ethics deeply unappealing, we hope that this chapter may at least help you to better understand what appeal *others* might see in the view.

However strong you judge the arguments in favor of utilitarianism to be, your ultimate verdict on the theory will also depend upon how well the view is able to counter [the influential objections that critics have raised against it](#).

The next chapter discusses theories of well-being, or what counts as being good for an individual.

Next Chapter: Theories of Well-Being

## How to Cite This Page

Chappell, R.Y. and Meissner, D. (2023). Arguments for Utilitarianism. In R.Y. Chappell, D. Meissner, and W. MacAskill (eds.), *An Introduction to Utilitarianism* <https://www.utilitarianism.net/arguments-for-utilitarianism>, accessed 3/13/2023.

## Resources and Further Reading

- Krister Bykvist (2010). *Utilitarianism: A Guide for the Perplexed*. Continuum.
- Robert Goodin (1995). *Utilitarianism as a Public Philosophy*. Cambridge University Press.
- Caspar Hare (2016). [Should We Wish Well to All?](#) *Philosophical Review*, 125(4): 451–472.
- Peter Singer & Katarzyna de Lazari-Radek (2017). Chapter 2: Justifications, in *Utilitarianism: A Very Short Introduction*. Oxford University Press.
- J.J.C. Smart (1973). An outline of a system of utilitarian ethics, in J.J.C. Smart & Bernard Williams, *Utilitarianism: For and Against*. Cambridge University Press.

- 
1. Daniels, N. (2020). [Reflective Equilibrium](#). *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.). ↩
  2. That is not to say that either answer is in fact equally good or correct, but just that you should expect it to be difficult to *persuade* those who respond to the conflicts in a different way than you do. ↩
  3. Of course, there may be exceptional circumstances in which stealing is overall beneficial and hence justified, for instance when stealing a loaf of bread is required to save a starving person's life. ↩
  4. Here it is important to consider the indirect costs of reducing social trust, in addition to the obvious direct costs to the victim. ↩
  5. Compare our defense of additive aggregationism in [another chapter](#), showing how, in practice, almost everyone endorses allowing sufficiently many small benefits to outweigh great costs to a few: “For example, allowing cars to drive fast on roads increases the number of people who die in accidents. Placing exceedingly low speed limits would save lives at the cost of

inconveniencing many drivers. Most people demonstrate an implicit commitment to aggregationism when they judge it worse to impose these many inconveniences for the sake of saving a few lives.”

See also Goodin, R. (1995). *Utilitarianism as a Public Philosophy*. Cambridge University Press.



6. Smart, J.J.C. (1956). Extreme and restricted utilitarianism. *The Philosophical Quarterly*, 6(25): 344–354.
7. Singer, P. (2011). Chapter 2: Equality and Its Implications, in *Practical Ethics* (3rd ed.). Cambridge University Press.
8. Preference utilitarianism accepts a [desire theory](#) of well-being.
9. The “veil of ignorance” thought experiment was originally developed by Vickrey and Harsanyi, though nowadays it is more often associated with John Rawls, who coined the term and tweaked the thought experiment to arrive at different conclusions. Specifically, Rawls appealed to a version in which you are additionally ignorant of the relative probabilities of ending up in various positions, in order to block the utilitarian implications and argue instead for a “maximin” position that gives lexical priority to raising the well-being of the worst-off. Vickrey, W. (1945). Measuring Marginal Utility by Reactions to Risk. *Econometrica*, 13(4): 329. Harsanyi, J.C. (1953). Cardinal Utility in Welfare Economics and in the Theory of Risk-taking. *Journal of Political Economy*, 61(5): 434–435. Rawls, J. (1971). *A Theory of Justice*. Belknap Press.
10. Harsanyi formalized his argument for utilitarianism in Harsanyi, J. (1978). [Bayesian Decision Theory and Utilitarian Ethics](#). *The American Economic Review*, 68(2): 223–228. For discussion about his proof, see Greaves, H. (2017). [A Reconsideration of the Harsanyi–Sen–Weymark Debate on Utilitarianism](#). *Utilitas*, 29(2): 175–213.
11. Singer, P. (2011). [The Expanding Circle: Ethics, Evolution, and Moral Progress](#). Princeton University Press.
12. Cf. Williams, E. G. (2015). [The Possibility of an Ongoing Moral Catastrophe](#). *Ethical Theory and Moral Practice*, 18(5): 971–982.
13. That is, fully informed moral agents who successfully do what they objectively ought to do. It would of course be no surprise if a misinformed or misguided moral agent made things worse.
14. Philippa Foot famously characterized the “spellbinding force” of utilitarianism (and consequentialism more broadly) as stemming from “the rather simple thought that it can

never be right to prefer a worse state of affairs to a better”., p. 198 Foot, P. (1985).

[Utilitarianism and the Virtues](#). *Mind*, 94(374): 196–209. ↩

15. The following arguments should also apply against virtue ethics approaches, if they yield non-consequentialist verdicts about what *acts* should be done. ↩
16. Absolutist deontologists hold such judgments to apply *no matter the consequences*. Moderate deontologists instead take the identified actions to be *presumptively* wrong, and not *easily* outweighed, but allow that this may be outweighed if a *sufficient* amount of value was on the line. So, for example, a moderate deontologist might allow that it is permissible to lie in order to save someone’s life, or to kill one innocent person in order to save a million. ↩
17. Samuel Scheffler noted that “either way, someone loses: some inviolable person is violated. Why isn’t it at least permissible to prevent the violation of five people by violating one?” Scheffler, S. (1984). *The Rejection of Consequentialism.*, p. 88 ↩
18. Scheffler, S. (1985). [Agent-Centred Restrictions, Rationality, and the Virtues](#). *Mind*, 94(375): 409–19. ↩
19. See, e.g., Chappell, T. (2011). [Intuition, System, and the “Paradox” of Deontology](#). In Jost, L. & Wuerth, J. (eds.), *Perfecting Virtue: New Essays on Kantian Ethics and Virtue Ethics*. Cambridge University Press., pp. 271–88. ↩
20. It is open to the deontologist to insist that it should be more important *to Jack*, even if not to anyone else. But this violates the appealing idea that the moral point of view is impartial, yielding verdicts that reasonable observers (and not just the agent themselves) could agree upon. ↩
21. For example, you might gaslight your spouse by remaining hidden in camouflage, when they could have sworn that you were just in the room with them. Or, as Foot (1978, 26) suggests, “An actor who fails to turn up for a performance will generally spoil it rather than allow it to be spoiled”. Foot, P. (1978). *The Problem of Abortion and the Doctrine of the Double Effect*. In *Virtues and Vices and Other Essays*. University of California Press. ↩
22. Beauchamp, T. (2020). *Justifying Physician-Assisted Deaths*. In LaFollette, H. (ed.), *Ethics in Practice: An Anthology* (5th ed.), pp. 78–85. ↩
23. Bennett, J. (1998). [The Act Itself](#). Oxford University Press. ↩
24. In a similar vein, Derek Parfit wrote that “Some of us ask how much of our wealth we rich people ought to give to these poorest people. But that question wrongly assumes that our wealth is ours to give. This wealth is legally ours. But these poorest people have much stronger moral claims to some of this wealth. We ought to transfer to these people (...) at least ten per

cent of what we earn”. Parfit, D. (2017). *On What Matters, Volume Three*. Oxford University Press., pp. 436–37 ↩

25. On the topic of sacrifice, John Stuart Mill wrote that “The utilitarian morality does recognise in human beings the power of sacrificing their own greatest good for the good of others. It only refuses to admit that the sacrifice is itself a good. A sacrifice which does not increase, or tend to increase, the sum total of happiness, it considers as wasted.” Mill, J. S. (1863). [Chapter 2: What Utilitarianism Is, Utilitarianism.](#) ↩
26. However, this does not mean that utilitarianism will strive for perfect equality in material outcomes or even well-being. Joshua Greene notes that “a world in which everyone gets the same outcome no matter what they do is an idle world in which people have little incentive to do anything. Thus, the way to maximize happiness is not to decree that everyone gets to be equally happy, but to encourage people to behave in ways that maximize happiness. When we measure our moral success, we count everyone’s happiness equally, but achieving success almost certainly involves inequality of both material wealth and happiness. Such inequality is not ideal, but it’s justified on the grounds that, without it, things would be worse overall. Greene, J. (2013). [Moral Tribes: Emotion, Reason, and the Gap Between Us and Them](#). The Penguin Press., p. 163; see also: [The Equality Objection to Utilitarianism.](#) ↩
27. In practice, the psychological phenomenon of *loss aversion* means that someone may feel *more upset* by what they perceive as a “loss” rather than a mere “failure to benefit”. Such negative feelings may further reduce their well-being, turning the judgment that “loss is worse” into something of a self-fulfilling prophecy. But this depends on contingent psychological phenomena generating *extra* harms; it is not that the loss is *in itself* worse. ↩
28. Bostrom, N. & Ord, T. (2006). [The Reversal Test: Eliminating Status Quo Bias in Applied Ethics.](#) *Ethics*, 116(4): 656–679. ↩
29. There are other types of debunking arguments not grounded in evolution. For instance, in the article [Objections to Utilitarianism](#) we describe another debunking argument: “consider that in most Western societies Christianity was the dominant religion for over one thousand years, which explains why moral intuitions grounded in Christian morality are still widespread. For instance, many devout Christians have strong moral intuitions about sexual intercourse, which non-Christians do not typically share, such as the intuition that it is wrong to have sex before marriage or that is wrong for two men to have sex. The discourse among academics in moral philosophy generally disregards such religiously-contingent moral intuitions. Many philosophers, including most utilitarians, would therefore not give much weight to the Christian’s intuitions about sexual intercourse.” ↩
30. de Lazari-Radek, K. & Singer, P. (2012). [The Objectivity of Ethics and the Unity of Practical Reason.](#) *Ethics*, 123(1): 9–31. ↩

31. Greene, J. (2007). [The secret joke of Kant's soul](#). In Sinnott-Armstrong, W. (ed.), *Moral Psychology, Vol. 3*. MIT Press. ↩