The Cluelessness Objection

Table of Contents

Utilitarianism directs us to promote overall well-being. But we cannot be certain how to do this. Worse, there are powerful reasons to think we are completely *clueless* about the long-run consequences of our actions, including whether they will be positive or negative overall. Does this make utilitarianism unworkable? Is it a reason to think that utilitarianism is *false*?

The Epistemic Objection to Consequentialism

James Lenman's Consequentialism and Cluelessness presents an influential *epistemic objection* against consequentialism (and hence, by extension, utilitarianism). We may reconstruct the argument roughly as follows:

> P1. We have no idea what the long-term effects of any of our actions will be.
>
> P2. But the long-term effects determine what we ought, according to consequentialism, to do. So, if consequentialism is true, we have no idea what we really ought to do—our reasons for action [1] lie beyond our epistemic grasp.
>
> P3. But an adequate ethical theory must be action-guiding—it cannot posit reasons beyond our epistemic grasp.
>
> Therefore,
>
> C. Consequentialism is not an adequate ethical theory.

Let us examine each of the three premises in turn.

Premise 1: Long-term Cluelessness

Imagine a doctor saving the life of a pregnant woman many centuries ago. [2] This seems like a clearly good act. Alas, it turns out that the woman was an ancestor of Hitler. So the seemingly-good act turned out to have actually-disastrous overall consequences.

This example illustrates how we might fail to grasp the long-term effects of our actions. But the point generalizes even to less dramatic actions, as small changes can ripple unpredictably into the future. For example, one's choice of whether or not to drive on a given day will "advance or delay the journeys of countless others, if only by a few seconds", [3] and they in turn will slightly affect others. Eventually the causal chain will (however slightly) affect the timing of a couple conceiving a child. A different sperm will fertilize the egg than would otherwise have been the case, leading to an entirely different child being born. This different person will make different life choices, impacting the timing of other couples' conceptions, and the identity of the children they produce, snowballing into an ever-more-different future. As a result, we should expect our everyday actions to have momentous—yet unpredictable—long-term consequences. Some of these effects will surely be very bad, and others very good. (We may cause some genocidal dictators to come into existence thousands of years from now, and prevent others.) And we've no idea how they will balance out.

Long-term consequences swamp short-term ones in total value. And because we generally can't predict the long-term consequences of our actions, it follows that we generally can't predict the *overall* consequences of our actions.

But there may be some exceptions. Proponents of longtermism believe that some actions—such as reducing existential risk—have robustly positive expected value in the long term. So, at a minimum, the sub-conclusion (premise 2) needs to be weakened to the claim that we've no idea what to do *other than work on reducing existential risks*. But even this weakened claim would remain surprising: it sure seems like we *also* have good reason to save lives in the here and now. The next section evaluates whether this is so.

Premise 2: Cluelessness and Expected Value

The natural response to cluelessness worries is to move to expectational consequentialism: promoting *expected* value rather than *actual* value. Further, as a multi-level theory, utilitarianism allows that we may best promote expected value by relying on heuristics rather than explicit calculation of the odds of literally every possible outcome. So if saving lives in the near term generally has positive expected value, that would suffice to defang the cluelessness objection.

Lenman distinguishes the "visible" (epistemically accessible) and "invisible" (entirely unknowable) consequences of an action. [4] Using this distinction, there is a very quick argument that saving lives has positive expected value. After all, if we've no idea what the long-term consequences of an action will be, then these "invisible" considerations are (given our evidence) simply *silent*—that is, speaking neither for nor against any particular option. So the visible reasons

win out, unopposed. For example, saving a child's life has an expected value of "+1 life saved", which doesn't change when our long-term ignorance is pointed out.

Lenman is unimpressed with this response,[5] but the reasons he offers are all highly disputable. Here we'll focus on his two primary objections.[6]

First, he suggests that expectational consequentialists must rely on controversial probabilistic indifference principles (the idea that, by default, we should assume that every possibility is equally probable).

In response, Hilary Greaves argues that some restricted principle of indifference seems clearly warranted in simple cluelessness cases, whatever problems might apply to a fully general such principle.[7] After all, it would seem entirely unwarranted to have asymmetric (rather than 50/50) expectations about whether saving an arbitrary person's life now was more likely to randomly cause or to prevent genocides from occurring millennia hence. So we can reasonably ignore such random causal factors.

However, as Greaves herself notes, this leaves open cases of "complex cluelessness" involving reasons to think that one option is *systematically better* than another for the long-term future, other reasons to judge the *opposite*, and it's unclear how to weigh the conflicting reasons against each other.[8] For example, if averting child deaths from malaria tends to result in a lastingly larger global population, you might think there are some reasons to judge this positively, and other reasons to judge this to be overall bad (due to "overpopulation"[9]). If we're confident of a systematic effect but just not sure of its direction, then it's less obvious that we can reasonably ignore it. At least, it doesn't seem like the principle of indifference appropriately applies here: it doesn't seem justified to assume that a larger population is equally likely to be good or bad.

But consequentialists may nonetheless repeat the earlier argument that "invisible" (unknowable) reasons can't guide our actions, so the only reasons left are the "visible" (knowable) ones, which speak in favor of saving lives and other seemingly-good acts until proven otherwise. This response does not rely on any principle of indifference. Instead, it stresses that the burden is on the skeptic to show *how* we should revise our initial judgment that saving a child's life has an expected value of "+1 life saved". It hardly seems better to throw up our hands in despair in the face of complex cluelessness. So until we're presented with a better alternative, it seems most reasonable to stick with our initial judgment.[10]

Second, Lenman assumes that, given the sheer immensity of the "invisible" long-term stakes, the "visible" reason for consequentialists to save a life must be extremely weak—merely "a drop in the ocean".[11] But this is mistaken. In absolute terms, saving a life is incredibly important. The presence of *even greater* invisible stakes doesn't change the absolute weight of this reason.

You might assume that, were consequentialism true, the strength of a reason for action must be proportionate to the action's likelihood of maximizing overall value. On this assumption, since the value of one life is vanishingly unlikely to tip the scales when comparing the long-term value of each option, saving one life must be an "extremely weak" reason to pick one option over another. But the earlier assumption is false. The strength of a consequentialist reason is given by its associated (expected) value in absolute terms: what matters is the size of the drop, not the size of the ocean.

Why Expected Value Matters

Specific objections aside, Lenman's larger worry is that it isn't clear *why* consequentialists should prefer prospects with greater expected value, if we're clueless about the actual consequences.[12]

This is a subtle issue. The point of being guided by expected value is *not* to increase our chance of doing the objectively best thing, as some risky prospects that are unlikely to turn out well may nonetheless be worth the risk.[13] Roughly speaking, it's a way to *promote value as best we can* given the information available to us (balancing stakes and probabilities).[14] After all, if there were an identifiably better alternative, to follow *it* would then maximize expected value. So if Lenman's critique were accurate, it would imply not that maximizing expected value is unmotivated, but rather that (contrary to initial appearances) saving a life lacks positive expected value after all.

If the question is instead asked, "Why think that saving a life has positive expected value?" then one may simply reply: "Why not? It's visibly positive, and invisible considerations can hardly be shown to count against it!"

Granted, cluelessness in the face of massive invisible long-term stakes can be angst-inducing. It should make us strongly wish for more information, and motivate us to pursue longtermist investigation if at all possible. But if no such investigations prove feasible, we should not mistake this residual feeling of angst for a reason to doubt that we can still be rationally guided by the smaller-scale considerations that we do see. To undermine the latter, it's not enough for the skeptic to gesture at the deep unknown. Unknowns, as such, are not epistemically undermining (greedily gobbling up all else that is known). To undermine an expected value verdict, you need to show that some alternative verdict is epistemically superior. Proponents of the epistemic objection, like radical skeptics in many other philosophical contexts, have not done this.[15]

Premise 3: The Possibility of Moral Cluelessness

The epistemic argument's final premise claims that *an adequate ethical theory must be action-guiding:* it cannot posit moral reasons beyond our epistemic grasp. But why think this? We may certainly *hope* for action-guidance. But if the world doesn't cooperate—if we're deprived of access to the morally-relevant facts—then it seems more appropriate to blame the world, not a moral theory that (rightly!) recognizes that unforeseeable events still matter.

Utilitarianism as a moral theory can be understood as combining (i) aggregative impartial welfarism as an account of the correct moral goals (i.e., what matters, or what we should care about), and (ii) the teleological principle that our reasons for action are given by applying *instrumental rationality* to the *correct moral goals*. This means that a misguided action must stem from either misguided moral goals or pursuing moral goals in an ineffective way.

A genuinely threatening objection to utilitarianism must then undermine one of these two sub-claims. Most commonly, critics challenge the utilitarian account of what matters, for example by suggesting that we should also care independently about rights, equality, or our nearest and dearest. But the cluelessness objection gives us no reason to doubt that future people genuinely matter, and hence that moral agents ought to care about the well-being of future people.[16] It may just be a sad fact about the world that we truly cannot know how to achieve our moral goals.

For example, suppose you must pull a magic lever either to the left or the right, and are told only that the fate of the world hangs on the lever's resulting position. You have no way of knowing which option will save the world. But it would be strange to conclude from this that the fate of the world does not morally matter. It would seem more reasonable to conclude that you're in a rough spot, and (in the absence of further evidence about which option is more likely to save the world) morality can offer you no useful guidance in these particular circumstances.

So premise 3 appears mistaken.[17] It's always possible that agents may be unable to know how to achieve their moral goals. In such a case, the true moral theory may fail to be action-guiding. But that does not undermine its truth. There's no principled reason to prefer an alternative theory that offers extra "guidance" without actually helping you to achieve the right moral goals.

All plausible theories should agree that overall consequences are *among* the considerations that matter (even if they diverge from consequentialism in claiming that other factors matter in addition). Moderate deontologists, for example, posit extra deontic constraints but allow that they may be overridden when the stakes are sufficiently high. This suggests that the cluelessness objection should be addressed to all moral theorists, not just consequentialists. These theorists may similarly reply that cluelessness is (at most) a practical difficulty, and not an objection to the *truth* of a moral theory.[18]

Conclusion

There is reason to doubt whether concerns about cluelessness really present an *objection* to utilitarianism at all. Cluelessness may just be a sad implication of the circumstances in which we find ourselves. But considerations of expected value, mediated via plausible heuristics, may continue to guide us nonetheless. We might reasonably take near-term expected value at face value, even if we've no idea about the long-term consequences of the acts in question. Moreover, even if long-term cluelessness swamps near-term expected value, there may still be some options

—like working to reduce existential risk—that have appreciably positive long-run expected value. So utilitarianism need not leave us entirely clueless about how to act, after all.

How to Cite This Page

Chappell, R.Y. (2023). The Cluelessness Objection. In R.Y. Chappell, D. Meissner, and W. MacAskill (eds.), <https://www.utilitarianism.net/objections-to-utilitarianism/cluelessness>, accessed 3/19/2024.

Resources and Further Reading

- Joanna Burch-Brown (2014). Clues for Consequentialists. *Utilitas*, 26(1): 105–119.

- Hilary Greaves (2016). Cluelessness. *Proceedings of the Aristotelian Society*, 116(3): 311–339.

- James Lenman (2000). Consequentialism and Cluelessness. *Philosophy and Public Affairs*, 29(4): 342–370.

- Andreas Mogensen (2021). Maximal Cluelessness. *The Philosophical Quarterly*, 71: 141–162.

- Andreas Mogensen & William MacAskill (2021). The Paralysis Argument. *Philosophers' Imprint* 21 (15): 1–17.

- David Thorstad, and Andreas Mogensen (2020). Heuristics for clueless agents: how to get away with ignoring what matters most in ordinary decision-making. *GPI Working Paper* 2-2020.

1. That is, the considerations that count in favor of acting one way rather than another. ↵

2. Adapted from James Lenman (2000). Consequentialism and Cluelessness. *Philosophy and Public Affairs*, 29(4): 342–370, p. 344. ↵

3. Hilary Greaves (2016). Cluelessness. *Proceedings of the Aristotelian Society*, 116(3): 311–339, p. 314. ↵

4. Lenman (2000), p. 363. ↵

5. See Lenman (2000), pp. 353–359. ↵

6. He offers four objections in all. The fourth presupposes the second, and so is addressed by our response to that. The third objects that we need to distinguish two very different reasons for

judging an act to lack expected value: (i) we might *know* that it makes no difference, or (ii) we might be *clueless* about whether it's incredibly good or incredibly bad. Given that these two epistemic states are so different, Lenman reasons, it makes no sense to treat them the same way.

It's true that this is a significant difference. But it's a mistake to assume that anything morally significant must change how we assess *acts*, when often *attitudes* are better suited to reflect such significance. We should feel vastly more angst and ambivalence—and strongly wish that more information was available—in a high-stakes "total uncertainty" case than in a "known zero" case. That seems sufficient to capture the difference. ↩

7. Greaves (2016), section IV. ↩

8. Greaves (2016), section V. See also Andreas Mogensen (2020). Maximal Cluelessness. *The Philosophical Quarterly*, 71: 141–162. ↩

9. For an exploration of whether the world is overpopulated or underpopulated, see Ord, T. (2014). Overpopulation or Underpopulation?, in Ian Goldin (ed.), *Is the World Full?*. Oxford: Oxford University Press. ↩

10. For a related defense of the "procedural rationality" of relying on heuristics in the face of cluelessness, see David Thorstad, and Andreas Mogensen (2020), Heuristics for clueless agents: how to get away with ignoring what matters most in ordinary decision-making. *GPI Working Paper 2, 2020.*

    Whether it's worth evaluating overpopulation concerns more deeply will depend on factors such as: (i) how many resources are at stake—more investigation is plausibly warranted for a grant-maker directing billions of dollars than for an individual donating a few hundred dollars; and (ii) how tractable the uncertainty seems, or what the expected value-of-information is of further investigation. For a small donor with little chance of swiftly resolving their uncertainty, it will often be most reasonable to entirely ignore complex cluelessness. ↩

11. Lenman (2000), p. 356. ↩

12. Lenman (2000), p. 360. ↩

13. For example, a 10% chance of saving a million lives is better in expectation than saving one life for certain, even though the latter option is 90% likely to yield a better outcome. The precise magnitudes and probabilities matter, not just whatever is "more likely" to be (however slightly) better. ↩

14. One important feature of maximizing expected value is that we cannot expect any subjectively-identifiable alternative to do better in the limit (that is, imagining like decisions being repeated a sufficient number of times, across different possible worlds if need be). ↩

15. One interesting proposal is that we should instead have *imprecise credences* covering a wide range of reasonable-seeming credal points. One worry with such proposals is that they may end up implying that we've just as much reason to support the *For Malaria Foundation* as the *Against Malaria Foundation*, which may undermine the apparent reasonableness of the starting assumptions. Cf. Andreas Mogensen (2021) [Maximal Cluelessness](). ↵

16. Strikingly, even Lenman (2000), p. 364, grants that "the invisible consequences of action very plausibly matter too," but he adds that "there is no clear reason to suppose this mattering to be a matter of moral significance any more than the consequences, visible or otherwise, of earthquakes or meteor impacts (although they may certainly matter enormously) need be matters of, in particular, moral concern. There is nothing particularly implausible here. It is simply to say, for example, that the crimes of Hitler, although they were a terrible thing, are not something we can sensibly raise in discussion of the moral failings or excellences of [someone who saved the life of Hitler's distant ancestor]."

   This is a strange use of "moral significance". Moral agents clearly ought to care about earthquakes, meteor strikes, and future genocidal dictators. (At a minimum, we ought to prefer that there be fewer of such things, as part of our beneficent concern for others generally.) An agent who was truly *indifferent* to these things would not be a virtuous agent: their indifference reveals a callous disregard for future people. So it could certainly constitute a "moral failing" to fail to care about such harmful events.

   On the other hand, if Lenman really just means to say that *which unforeseeable consequences actually occur* shouldn't affect our assessment of a person's "moral failings or excellences", then this seems a truism that in no way threatens consequentialism. It's a familiar point that many forms of agential assessment (e.g. rationality, virtue, etc.) are "internalist"—supervening on the intrinsic properties of the agent, and not what happens in the external world, beyond their control. [Hybrid utilitarians]() comfortably combine such internalism about agential assessments with a utilitarian account of our reasons for action. ↵

17. Unless interpreted in such a way as to no longer demand guidance where none is possible. While it's surely fine to have fact-relative reasons that outstrip our epistemic grasp, a more compelling version of the premise might just claim that evidence-relative reasons must be within our epistemic grasp. But then it risks collapsing into mere tautology: by definition, the "evidence-relative reasons" posited by any theory—including consequentialism—will be epistemically accessible (assuming that "evidence" and "epistemic access" go together). The question instead becomes what (if any) evidence-relative reasons the theory implies that we have. ↵

18. Though there's some reason to think that the practical difficulties may be even worse for non-consequentialists. See Andreas Mogensen & William MacAskill (2021). [The Paralysis Argument](). *Philosophers' Imprint* 21 (15): 1–17. ↵