Ethic Theory Moral Prac (2015) 18:759–766
DOI 10.1007/s10677-015-9602-8

The Unreliability of Foreseeable Consequences: A Return to the Epistemic Objection

Samuel Elgin[1]

Accepted: 1 April 2015 / Published online: 5 May 2015

Abstract Consequentialists maintain that an act is morally right just in case it produces the best consequences of any available alternative. Because agents are ignorant about some of their acts' consequences, they cannot be certain about which alternative is best. Kagan (1998) contends that it is reasonable to assume that unforeseen good and bad consequences roughly balance out and can be largely disregarded. A statistical argument demonstrates that Kagan's assumption is almost always false. An act's foreseeable consequences are an extremely poor indicator of the goodness of its overall consequences. Acting based on foreseeable consequences is barely more reliably good than acting completely at random.

Keywords Consequentialism · Epistemic objection · Statistics

"*I foresee all sorts of unforeseen problems*"—Sir Humphrey (Yes, Prime Minister)

1 Introduction

In 1895 Thomas Austin released 24 rabbits in Australia, intending to hunt them for sport. The initial consequences were roughly as foreseen—a few hunters enjoyed themselves and a few rabbits did not. But, having no natural predators, the rabbits multiplied. Within a decade millions roamed the continent with disastrous economic and ecological consequences. The results of Austin's act were far worse than predicted. This is a classic example of the significance of unforeseen consequences. Although we are rarely able to track an act's long-term consequences so precisely, everyone is familiar with situations that do not turn out as expected.

✉ Samuel Elgin
samuel.elgin@yale.edu

[1] Yale University, New Haven, CT, USA

Objective consequentialism contends that an action is morally right just in case it has the best consequences of any action available to an agent at a given time. Although consequentialists diverge over what constitutes the best consequences, all face a potentially devastating epistemic objection: we do not know what the consequences of our actions will be. An agent may have excellent reasons to expect her action to have good consequences when an unanticipated disaster results. In this case, the consequentialist maintains that her act is morally wrong. Of course, the opposite can also occur. An agent might reasonably expect her action to have bad consequences when something unanticipated and fortuitous results. The consequentialist holds that, in this case, the agent's act is morally right. Because agents are unable to know what the consequences of their actions will be, the epistemic objection maintains, act consequentialism affords no insight into how to act. Unsurprisingly, consequentialists disagree. As Kagan says:

> The problem of uncertainty may not be incapacitating. After all, life is full of risks and uncertainties—not just in moral cases, but everywhere. Although we may lack crystal balls, we are not utterly in the dark as to what the effects of our actions will be. We are able to make reasonable, educated guesses. And thus we can—and do—set ourselves goals and choose our acts with an eye toward how we are most likely to promote those goals. Uncertainty need not lead to paralysis. (Of course, it remains true that there will always be a very small chance of some totally unforeseen disaster resulting from your act. But it seems equally true that there will be a corresponding very small chance of your act resulting in something fantastically wonderful, although totally unforeseen. If there is indeed no reason to expect either, then the two possibilities will cancel each other out as we try to decide how to act.) (1998:64).

In most situations, Kagan argues, agents can foresee many consequences of their actions. Still, mistakes happen. Fallibilism with respect to the best consequences is reasonable. Skepticism is not.

The problem with Kagan's defense is that the basis of the epistemic objection does not rest on rare consequences of extreme significance. Rather, it rests on the fact that the long-term unforeseeable consequences vastly outnumber the foreseeable ones, and we are clueless about what the long-term consequences will be. Lenman (2000) has made this argument using persuasive examples. I argue that the success of these examples is no fluke. Nor is cluelessness, as he intimates, primarily a problem for momentous acts like procreation and killing. There is a strong statistical reason to deny that unforeseeable consequences will cancel out as Kagan thinks. As the number of unforeseeable consequences increases, the probability that the foreseeable consequences are representative of the overall consequences diminishes. In choosing between two possible courses of action, picking the action with the best foreseeable consequences is only slightly more reliable than flipping a coin.

2 The Informal Gloss

Suppose Rob finds a coin and decides to give it a flip. Assuming the coin is fair, there is a 50 % chance it will come up heads and a 50 % chance it will come up tails. If he flips it twice there are three possibilities: a 25 % chance it will come up heads both times, a 50 % chance it will come up heads once and come up tails once, and a 25 % chance it will come up tails both times. If he continues flipping the coin, the number of possible outcomes increases. Solutions to binomial equations specify the odds that the coin will come up n times given m total flips.

Binomial distributions have surprising implications. Consider, for example, the odds that Rob's coin will come up heads at least five more times than it comes up tails. If there are fewer than five flips, the odds of this happening are 0 %. Clearly, if he flips the coin only four times, it is impossible for it to come up heads five more times than it comes up tails. If he were to flip the coin exactly five times there is a 3.125 % chance it would come up heads five times in a row. As the number of flips increases, the odds that the coin will come up heads at least five more times than it comes up tails rises—approaching 50 % as the number of flips approaches infinity.

This is counterintuitive. Everyone knows that as the number flips increases, the average number of heads gets closer to 50 %. This might seem incompatible with the claim that the odds that the coin will come up heads at least five more times than it comes up tails also rises. However, what counts as 'close to 50 %' changes as the number of flips increases. If Rob were to flip his coin seven times and it came up heads six out of the seven, it is not the case that close to half of the flips came up heads. But if he were to flip the coin 1001 times and it came up heads only five more times than it came up tails, the number of heads would be very close to half the total number of flips. With that many trials, five flips one way or the other is relatively insignificant.

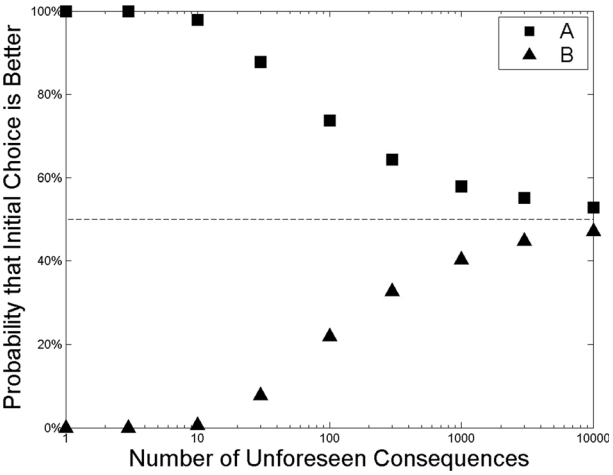
Obviously, binomial equations are not merely concerned with coin flips. They describe probable outcomes of many sorts of repeated events. Here, I use them to determine how unforeseen consequences are likely to aggregate. If there are relatively few unforeseen consequences, the foreseen ones are a reliable guide to what the overall consequences will be. But as the number of unforeseen consequences increases, the foreseen ones become less and less significant—like knowing that a coin will come up heads 10 times in 10,000 flips and being ignorant of the rest. If objective consequentialism is correct, we are in an extremely bad situation with regard to epistemically determining the right thing to do.

3 The Formal Argument

Let a *utile* be one positive unit of utility. Extremely good consequences yield many positive utiles, neutral consequences yield none, extremely bad consequences yield many negative utiles, etc. Suppose agent *S* is debating between two possible courses of action—call them *A* and *B*. *S* is not completely ignorant. She can foresee some, but not all, of the consequences of each action. Based on the foreseeable consequences, she predicts that *A* will result in 10 utiles while *B* will result in 0. What is the probability that performing *A* would actually have better overall consequences than performing *B* would? If the probability is close to 50 %, *S*'s knowledge of the foreseeable consequences is an unreliable guide. Nearly half the time, acting on the best foreseeable consequences yields a worse outcome than an available alternative.

To show this, I assume that there is a 50 % chance that an unforeseen consequence will be good and a 50 % chance that it will be bad (since the consequences are unforeseen, there is no reason to privilege one sort over the other). Morally neutral consequences have no net effect and can be ignored. So I assign 50 % to the probability that an unforeseen consequence will increase an action's overall utility and a 50 % chance that it will decrease the action's overall utility. To simplify the math, I assume that each unforeseen consequence has an equal measure of positive or negative utility. Each good consequence increases the total utility by 1 and each bad consequence decreases the total utility by 1. So, on this model, the utility of each unforeseen consequence is one tenth the foreseen difference in utility.

I model this situation as the binomial distribution of consequences for actions *A* and *B*, where *A* has an initial utility of 10 and *B* has an initial utility of 0. If there are 5 or fewer unforeseen consequences, the odds that *B* will be better than *A* are 0 %. If there are 6 or more unforeseen consequences, the odds are higher. The following graph shows the odds that each action will result in better overall consequences for various numbers of unforeseen consequences. The equations I used to generate this graph appear in the Technical Appendix.[1]



As the number of unforeseen consequences increases, the odds that *A* will have better overall results diminishes. If there are more than 1000 unforeseen consequences there is over a 40 % chance that performing *B* will have better consequences than performing action *A* will. According to the consequentialist, this means that there is a 40 % chance that performing *B* is the morally right thing to do. As the number increases further, the probability approaches 50 %. With sufficiently many unforeseen consequences, foreseen consequences are an extremely unreliable guide as to the best outcome.

Other moral theories face analogous epistemic problems. Absolute deontology—the theory that agents must abide by certain rules, and otherwise should act so as to maximize the best consequences—faces the same problem as objective consequentialism in cases in which the rules do not apply.[2] Moderate deontology—the theory that agents must abide by certain rules unless the consequences of following those rules are bad enough, and otherwise should act so as to maximize the best consequences—faces the problem twice over. They are ignorant of whether the consequences are 'bad enough' in cases in which the rules apply, and ignorant of what the best consequences are in cases in which they do not.

Some subjectivist theories have the resources to evade the worry. Naïve subjective consequentialism—the theory that an agent's act is morally right just in case she believes that it will have the best consequences of any available alternative—is unaffected. Agents' beliefs are frequently incorrect, but this does not entail that agents are ignorant about the content of their beliefs. A refined form of subjective consequentialism—which holds that an agent's act is

[1] I use a logarithmic scale because of the large numbers of consequences that must be surveyed to show the trend.
[2] Versions of absolute deontology where the rules depend on how bad the consequences of an action are presumably also face the epistemic problem in cases in which the rules apply.

morally right just in case her evidence most strongly points in favor that that act will have the best consequences of any available alternative—is similarly untouched.

Other subjectivist theories face a far deeper problem. Consider reasonable belief consequentialism, which holds that an agent's act is morally right just in case she could reasonably believe that it would have the best consequences of any available alternative. Arguably, beliefs of the form 'φ will have the best consequences' are almost always unreasonable. At best there is only slightly greater than a 50 % chance that such beliefs are correct. Reasonable agents remain agnostic about which actions will have the best consequences. But if no acts can be reasonably believed to have the best consequences then (according to reasonable belief consequentialism) there are no acts that are morally correct.

4 Consequentialist Responses

Consequentialists (and other affected moral theorists) could respond to this argument in several ways. Each objection warrants careful consideration, but all are unsuccessful. The following are the most plausible objections that they could raise:

Unforeseen consequences are not uniformly significant. The calculation was based on the assumption that all unforeseen consequences had equal measures of utility. This is preposterous! Some consequences (e.g., that a surgery was successful) are extremely good. Others (e.g., that a pencil point did not break) are only marginally good. Surely it is mistaken to assign them equal weight.

This is so. In the actual world, not all unforeseen consequences are equally significant. We could accommodate this by randomly assigning different weights to different unforeseen consequences. Two points are worth mentioning. Firstly, any particular distribution of weights would need just as much justification as an equal distribution—and it is not obvious what such a justification would be. Secondly, although this would significantly complicate the math needed to calculate the odds that a given action would have the best consequences, there is no reason to think that it would result in the foreseeable consequences reliably indicating the overall utility of an action. This simplifying assumption is one that enables manageable calculation while not being responsible for the overall trend in utility aggregation.

The unforeseen consequences of A are more likely to be good than those of B. A has better foreseeable consequences than *B* does. Inductively, shouldn't we assume that the unforeseen consequences of *A* are also more likely to be good than those of *B*? If so, the calculation was flawed. It assumed that each action was equally likely to have good and bad unforeseen consequences.

If justified, this is a powerful objection. The odds that *A* will have the best overall consequences are high if the unforeseen consequences of *A* are significantly more likely to be good than those of *B*. But does induction justify such an assumption? Often, acts' unforeseen consequences are extremely dissimilar from their foreseen ones. For example, if someone decides to wear a red shirt to show solidarity with her local sports team, the unforeseen consequences of wearing red might be completely unrelated to those she predicts.

And there is no good reason to think that the unforeseen consequences of wearing blue are likely to be worse than those of wearing red. It is unwarranted to assume that the unforeseen consequences of *A* are more likely to be good than those of *B*.

There are insufficiently many unforeseen consequences for our ignorance to be problematic. If the effects of a given act on all subjects with welfares are suitably short-lived, unforeseen consequences are unlikely to swamp the foreseen ones. Even if the unforeseen consequences were ten times as significant as those we can foresee, there is over a 78 % chance that *A* would have the best results.[3]

Shortly before the extinction of the last welfare-bearing creatures, the foreseeable consequences of an act may well be a good guide to its overall consequences. This is an exceptional circumstance. Acts standardly have vast numbers of consequences that affect welfare and that we quickly lose track of. Rather than disappearing like ripples in a pond (Smart 1973), the consequences of our actions persist—as, for example, damage to the ozone remains even if we have lost track of each original source. The conviction that the morally significant consequences of a seemingly trivial act are short-lived is more than dubious. We have good reasons to think that it is false.

Temporally distant consequences count less than temporally close ones. Unlike previous objections, this response constitutes a change to the consequentialist's thesis. Rather than weighting all consequences equally, she takes the short-term consequences to be particularly significant. There are two ways that such a theory could be fleshed out.

First, she could select a time *t* in the future and claim that consequences that occur before *t* count morally, while those that occur after *t* do not. If we are close enough to *t*, our ignorance of long-term consequences is morally irrelevant. Although this avoids the epistemic objection, it faces other serious problems. What privileges *t* over other times in the future? Even in principle, how could *t* be privileged over a split-second later? Such a theory also diagnoses certain cases unintuitively. Someone who sets a timer on a bomb to go off immediately before time *t* is morally culpable for the deaths that result. Someone who merely sets the timer for two seconds later is morally in the clear.

Second, she could argue that future consequences diminish gradually in weight over time. Consequences that occur in the near future are particularly morally significant. Ones that occur in the distant future still count, but count substantially less than short-term consequences do. If future consequences decrease in weight quickly enough, an act's foreseeable consequences reliably indicate whether or not it is morally correct. Such a consequentialist need not justify selecting a particular time *t* after which consequences change from counting completely to not counting at all. Still, such a theory diagnoses certain cases unintuitively. Suppose two people plant bombs underneath crowded areas in London. Both confidently and reasonably expect that when their bombs detonate, many innocent people will die. However, the first person sets the timer in her bomb for 1 month while the second sets the timer for 10 years. Intuitively the second is as morally reprehensible as the first. The fact that a substantial amount of time passes before the negative consequences of her act arise does not mitigate her moral responsibility at all.

Causally distant consequences count less than causally proximate ones. A consequentialist might change tack. In the cases mentioned above, there was a significant difference in time delay. But the causal chains linking the agents' setting the timers to the lethal explosions were both fairly short. Perhaps consequences that occur farther along a causal chain—rather than those that occur farther in the future—should receive less weight. If causally distant consequences decrease in weight quickly enough we need not be skeptical about which act will be best.

Setting aside the difficult task of differentiating links in a causal chain, this too has unintuitive results. Consider a machine designed by Rube Goldberg. This machine has a large and unnecessary number of links in a causal chain as it completes a simple task. Still, it is

[3] See Technical Appendix.

effective and people can often figure out precisely what it will accomplish. If the number of links in a causal chain were morally significant, committing murder with this machine would be morally preferable to using a knife. After all, the machine has far more links in its causal chain than the knife does. This is unintuitive.

Of course, sometimes agents seem not to be morally responsible for distant (either temporally or causally) consequences of their actions. If Jane's act of raising her hand triggers a hurricane via the butterfly effect, intuitively she is not responsible for the damage that results. But her innocence does not arise from the distance of the consequences. Rather, she is innocent because it is hopelessly difficult to predict consequences like that. If Jane had good reasons to expect that lifting her hand would cause a hurricane, we would take her to be culpable after all. These are precisely the sorts of considerations that push some philosophers toward more subjectivist approaches.

5 Conclusion

Moral skepticism results from objective consequentialism. Rarely, if ever, is an agent in a position to know (or even to reasonably believe) that a given course of action will have the best overall consequences. It is overwhelmingly likely that most of our acts have consequences that affect welfare—far more than those we can predict. In such cases, regardless of how many consequences we foresee, we have no reliable method of determining which acts will have the best results. If the morally right action is the one with the best results, we are nearly blind in determining what we should do. Consequentialism requires particular action in spite of profound ignorance.

Acknowledgments I would like to thank Shelly Kagan, Catherine Elgin and the attendees of the 2014 BSET conference for providing comments on earlier versions of this paper.

Technical Appendix

We model the choice situation as a binomial distribution and compare the various probabilities of consequences resulting from actions *A* and *B*. Situations in which the action with the best consequences are those that result from *A* are those in which the net utility of selecting *A* minus the net utility of selecting *B* is greater than 0.

Let *n* be the number of unforeseen consequences. The probability (for both actions *A* and *B*) that there will be *k* beneficial unforeseen consequences is determined by the following:

$$P_n^k = \frac{n!}{2^n k!(n-k)!}$$

As the number of unforeseen consequences increases, the factorials become extremely large. One way to calculate them is with the following recursive function.

$$P_n^0 = \frac{1}{2^n}$$

$$P_n^k = P_n^{k-1}\left(\frac{n-k+1}{k}\right)$$

Let V_a and P_a represent distinct net utile values and the probability that that net utile value will occur respectively for action A. Similarly, let V_b and P_b represent distinct net utile values and the probability that that net utile value will occur respectively for action B. The overall probability Q that action B will have a greater utility than action A will is given by the following:

$$Q = \sum_{k_1=0}^{n} \sum_{k_2=0}^{n} \left[\begin{array}{c} Pa_{k_1} \cdot Pb_{k_2} (\text{if } Vb_{k_2} > Va_{k_1}) \\ 0 (\text{otherwise}) \end{array} \right]$$

for n unforeseen consequences. Due to the difficulty in calculating binomials for $n>1000$, it is useful to use a normal approximation. The results of the binomial probability and the normal approximation are given in the table below:

n	Binomial Probability	Normal Probability
20	0.04035	0.05692
50	0.13563	0.15866
100	0.21838	0.23975
200	0.29119	0.30854
500	0.36399	0.37591
1000	0.40286	0.41153

If there are 1000 unforeseen consequence there is a greater than 40 % chance that action B has better consequences than action A does. The fact that the foreseeable consequences are such that A has 10 more utiles than B does is hardly a reliable indication that A will be better. And as the number of unforeseen consequences passes 1000, the odds that A is the best course of action decrease still further—approaching 50 %.

References

Kagan S (1998) Normative ethics. Westview Press, Boulder

Lenman J (2000) Consequentialism and cluelessness. Philosophy & Public Affairs 29:342–370

Smart JJC (1973) An outline of a system of utilitarian ethics. In: Smart JJ, Williams B (eds) Utilitarianism—for and against. Cambridge University Press, Cambridge