

# Debunking Objective Consequentialism: The Challenge of Knowledge-Centric Anti-Luck Epistemology

Paul Silva Jr.  
University of Cologne

It is a part of our commonsense perspective on the world that we know and have some justified beliefs about the moral status of our prospective actions. With these we also attempt to make our way towards a theoretical understanding of what makes these moral beliefs true, i.e. we attempt to home in on the correct normative theory of ethics. It is generally assumed that the search for the correct normative theory of ethics is not absolutely futile, i.e. at the very least we can find reasons to justifiedly *increase* to some degree our confidence in some normative theories over others. Certain versions of objective act consequentialism are widely thought to be among the normative theories of ethics that we should take most seriously and thus be comparatively more certain about. In a large part, this is due to its ability to explain important parts of our commonsense moral perspective.

In what follows I will explain why this is mistaken from the perspective of knowledge-centric anti-luck epistemology. For according to it there are modal anti-luck demands on both knowledge and justification, and it turns out that our beliefs about the moral status of our prospective actions are almost never able to satisfy these demands if objective act consequentialism is true. Accordingly, if objective act consequentialism is true, we neither know nor have justified beliefs about the moral status of our prospective actions. So a problematic kind of applied moral skepticism obtains. As I will explain, this kind of applied moral skepticism introduces problematic limits on our ability to use objective act consequentialism's explanatory power to justifiedly increase our confidence in its truth. This is, in part, a product of higher-order defeat as I explain in the final section. There is, however, a silver lining for objective act consequentialists. For there is at least one type of objective act consequentialism, prior existence consequentialism, that is poised to avoid at least some of the epistemic problems discussed in this paper.

## 1 Knowledge-Centric Anti-Luck Epistemology and Justificatory Defeat

Take a standard lottery case:

### LOTTERY

I have a ticket for a fair lottery with very long odds. The lottery has been drawn, although I have not heard the result yet. Reflecting on the odds involved I conclude that (L) *my ticket is a loser*. Besides my (accurate) assessment of the odds, I have no other reason to think my ticket is a loser. As it turns out, my belief that I own a losing ticket is true.

It is widely thought that L cannot be known in these circumstances. The explanation for the unknowability of L in these circumstances is not that I lack strong evidence in favor of L. L is extremely probable on my evidence. Rather, the unknowability of L is thought to be owed to the fact that my belief, even if true, would be in some sense true by “epistemic luck”—a kind of luck that is incompatible with knowing. Generally, the nature of this epistemically problematic form of luck is thought to involve a *modal defect* in my believing L in these circumstances. The two leading accounts of this modal defect locate the problem with my belief not at the actual world where my belief is true, but in a relation I bear toward myself in nearby possible worlds where I falsely believe

L. The sensitivity account and the safety account are the leading accounts of that relation. Here is a common formulation of these proposed requirements on knowledge:

#### SENSITIVITY

Had it been that P, S would not still have believed that P (had S used the method that S actually used to determine whether P).

#### SAFETY

S could not have easily had a false belief as to whether or not P (using the method that S actually used to determine whether or not P).<sup>1</sup>

If either principle is correct, then L cannot be known in Lottery for neither principle is satisfied in that case. It is easy to see why Sensitivity fails in Lottery. My belief in L—that my ticket is a loser—is based solely on my statistical evidence. But I could have that same evidence even if my belief were false, i.e. even if there is some nearby world where my ticket is a winner. Because of this, I would still believe L on the basis of that same statistical evidence even were L false. So Sensitivity offers a straightforward explanation of the idea that L is unknowable in Lottery.

It is a bit trickier to see how Safety explains the same fact. The idea that a belief *could not have easily been false* is the idea that the world would have to be significantly different in order for the target to belief to be false. For example, I truly believe that I will not become a US Navy Seal this year. And this belief could not have easily been false in the sense that in order for it to be false the world would have to be quite different in a variety of respects: the relevant governing bodies of the US Navy would have to decide that it's advisable to lower their fitness standards (dramatically) in order for me to be considered admissible as a Seal, or there would have to be some kind of conspiracy where all relevant individuals involved in assessing Seal candidates willingly lie about my admissibility, or something else equally unusual would have to happen. Given my age, my limited physical fitness, my limited ability to persevere through physical pain, my lack of political power to orchestrate a conspiracy, and my intention not to join the Seals, etc. it could not have easily been false that I will not become a Seal this year. So this belief satisfies Safety, and so is a candidate for knowledge. In contrast, as Pritchard (2008: section 4) explains, my belief in L in Lottery could easily have been false since “all that need be different in order for one's ticket to be a winning ticket is that a few numbered balls fall in a slightly different configuration.” Intuitively, such worlds are not significantly different from the actual world and so my belief could not have easily been false.<sup>2</sup>

While the unknowability of L in Lottery is widely held, some hold out hope that L might never the less be *justifiedly believable* in Lottery (e.g. standard Lockean evidentialists about justification). But knowledge-centric theorists have generally argued for theses about the relation between knowledge and justification that are inconsistent with this view. While I cannot here go into the details, the general motivation connecting knowledge and justification has to do with the normative role of knowledge in our assessment of belief. First, not only does one intuitively fail to know in Lottery cases, one intuitively *shouldn't* hold, and so can't justifiedly hold, a lottery belief. If justification requires knowledge, this fact about lottery beliefs can easily be explained. Second, the idea that justification requires knowledge can explain why certain Moorean beliefs shouldn't be held, e.g. I believe P, but I do not know P. Third, a knowledge requirement on justification can explain our critical practices of assessing others' beliefs. For instance, if someone believes P, it's perfectly normal to object to their believing P on the basis of the fact that they don't know it. A knowledge requirement on justification would also explain this.

---

<sup>1</sup> Pritchard (2005; 2008; 2012).

<sup>2</sup> See Rabinowitz (2019) for further discussion of safety conditions.

It turns out there are different ways of putting knowledge at the center of the justification of belief. For example, some knowledge-centric theorists think that knowledge is necessary and sufficient (and to be token-identified with) justified belief (Williamson 2013; Sutton 2007; Littlejohn 2012):

$J \leftrightarrow K$ : S justifiably believes that P iff S knows that P.

If correct, then not only can L not be known in Lottery, it cannot be justifiably believed in Lottery either. Alternatively, some knowledge-centric theorists have argued that justification requires something just shy of knowing: *being in a position* to know (Bird 2007; Ichikawa 2014; Rosenkranz 2018):

$J \rightarrow PK$ : S justifiably believes P only if S is in a position to know P.

If correct, the impossibility of knowing L entails the impossibility of justifiably believing L in Lottery also.<sup>3</sup>

Now, if it's possible to have justified *false* beliefs, then perhaps it's possible to fail to know L while never the less having justification for certain higher-order beliefs: the belief that one knows L (though one doesn't) or the belief that one is justified in believing L (though one isn't). But notice both  $J \leftrightarrow K$  and  $J \rightarrow PK$  entail the defeat of these higher-order beliefs. For if either  $J \leftrightarrow K$  or  $J \rightarrow PK$  are true, it's *impossible* to have justified false beliefs if they cannot ever constitute knowledge. Accordingly, it's not just first-order lottery beliefs that are unjustified, these higher-order beliefs are also unjustified. So justification for both first-order and higher-order beliefs are lost.

Unsurprisingly, the idea that knowledge must be a bare possibility for one to have justification has seemed to some a difficult bullet to bite. Accordingly, some have sought to unify the intuitions driving knowledge-centric views of justification without committing themselves to the idea that knowledge or possible knowledge is required for justified belief. For example, Smithies (2012) has argued that justification for the belief that P requires that one enjoy justification to believe a higher-order claim about the knowability of P, i.e. that one have justification to believe that one is in a position to know P:

$J \rightarrow JPK$ : S justifiably believes P only if S has justification to believe that she's in a position to know P.

If this is correct then there is room for both first-order and higher-order justified beliefs in L. But of course such justified beliefs are limited to those who are *sufficiently ignorant* of the fact that one cannot know L in Lottery, i.e. those who lack access to reasons sufficient to defeat the claim that one *can* know L in Lottery. Accordingly,  $J \rightarrow JPK$  leaves many of us who are reflective about lottery cases in much the same position as  $J \leftrightarrow K$  and  $J \rightarrow PK$ : we cannot know L, we cannot

---

<sup>3</sup> Alternatively, some knowledge-centric theorists have argued that justification should be understood virtue-theoretically in terms of exercises of knowledge-yielding abilities, competences, or dispositions (Miracchi 2015; Silva 2017; Kelp 2018; Millar 2019):

$J \rightarrow EKD$ : S justifiably believe P only if S's belief is produced from an exercise of a knowledge-yielding ability (process, competence, disposition).

Provided such competences are understood in an anti-luck fashion as requiring safety or sensitivity, then we will get the same result here as with  $J \rightarrow PK$ . However, Silva, Miracchi, and Kelp all seem to allow for justified beliefs in standard gettier cases (where Safety and Sensitivity are not satisfied) and thus do not require for the exercise of those abilities that one believe safely or sensitively. Millar (2019) may be an exception to this.

justifiedly believe L, and we cannot even falsely justifiedly believe that we have justification to believe L.

As formulated, each of the knowledge-centric principles of justification mentioned above concern only doxastic justification. It is typically important to keep in mind the difference between propositional justification (=having justification to  $\phi$ ) and doxastic justification (=justifiedly  $\phi$ -ing). I will not make much of this in what follows and switch freely between the two locutions. This will make no difference under the assumption that having propositional justification requires at least the bare possibility of doxastic justification. Since knowledge of lottery beliefs is impossible, this will make both doxastic and propositional justification impossible to come by on the knowledge-centric views mentioned above.

In what follows I'll explain the surprising problem that anti-luck epistemology and knowledge-centric epistemology generate for ethical consequentialists.

## 2 Objective Act Consequentialism

For present purposes let a *consequentialist* about moral requirements be anyone who thinks that the truth of claims about what actions are morally required (forbidden, optional) depend solely on the long term consequences of that action in the following way:

### CONSEQUENTIALISM

S is *required* to perform action A if the the long term net value of A-ing is greater than the long term net value of performing any alternative action. If the the long term net value of A-ing is less than the the long term net value of performing some alternative action, then A-ing is *wrong*. If the the long term net value of A-ing is the same as the the long term net value of some alternative and there is no other alternative action with a higher net value, then A-ing is *optional*.

### NET VALUE

The net value of an action is the amount of value that results from that action in a specified period of time minus the disvalue that results from that action in that same period of time.

This leaves open a number of dimensions along which to specify one's preferred version of (objective, act) consequentialism. One could take an egoist view on which all that matters is how one's present actions maximize the net amount of pleasure of one's experiences in their life—in which case 'the long term' is just the duration of one's life. Alternatively, one could take a classical utilitarian view on which right action is determined by actions maximizing the net pleasure for all sentient beings who stand to be impacted by one's actions—in which case 'the long term' is just the duration of the effects of one's potential actions. This could be 10 minutes from the time of action, 10 years, or 10,000 millennia. Alternatively, one could take a form of prior existence utilitarianism where all that matters for fixing right action is how one's prospective actions would impact people who already exist (our would come to exist irrespective of which action is performed)—in which case 'the long term' is limited to the lifespan of those individuals.<sup>4</sup> While the forms of consequentialism I'm explicitly discussing are the maximizing varieties, satisficing varieties will have an equally difficult time avoiding the epistemic problem I develop for maximizing varieties.<sup>5</sup> Also, while consequentialist views suffer perhaps the worst from

---

<sup>4</sup> For critical discussion of prior existence utilitarianism see Singer (2011).

<sup>5</sup> If Sensitivity is right, this will be because our view as to whether or not a prospective action satisfies net value will still have to rely on something like premise (1) of the CMR (see below). While if Safety is right, this will be

knowledge-centric anti-luck epistemology, any moral theory that creates space for the moral status of at least *some* actions to be determined by the net value of their consequences will face a version of this problem.<sup>6</sup>

Let's start by considering *high-stakes moral beliefs* about the actual world. The belief that it's actually wrong to murder the entire Rohingya population of Myanmar, the belief that it's actually wrong to steal large sums of money from effective charities that would use it to alleviate the suffering of many, the belief that it's actually wrong to euthanize every other baby in the world, the belief that it's actually wrong to drop an atomic bomb near a large population to observe its negative effects across that population, etc. These are all claims about prospective actions in the actual world as opposed to merely possible worlds. When it comes to merely possible worlds, we can often *a priori* specify the net value of consequences of our potential actions more or less arbitrarily. Not so with the actual world. What I'm concerned with is the epistemic standing of our moral beliefs about actions we and others can perform in the *actual* world.

One thing these moral beliefs have in common is that they concern *foreseeably* high-stakes actions, i.e. these are prospective actions that are specified in such a way that it is assumed that the individual for whom they are prospective actions is in a position to know that *in the near term* an exceptionally bad moral state of affairs would result from performing them. Now, there can be prospective high-stakes actions in circumstances where every alternative action is also a high-stakes action (think of trolley cases where the numbers on the tracks are roughly equal). These are not the kind of circumstances I have in mind in what follows. Our moral beliefs about these kinds of high-stakes actions will be controversial and skepticism about them will be far less troubling. Rather, I'm limiting reflection to cases where a prospective high-stakes action has at least one prospective low-stakes alternative that one could easily perform. For example, a military leader's ability (a) to murder or displace all, or nearly all, of Myanmar's Rohingya population as well as his ability (b) to not murder or displace any of them. Ordinary moral judgements would affirm that performing (a) is morally wrong when—though perhaps not only when—(b) is an available prospective course of action.

High-stakes moral beliefs are common, generally uncontroversial, and often function as starting points (and sometimes as fixed points) in non-skeptical moral theorizing. Their evidential usefulness in moral theorizing is owed to the fact that they seem to be justified and knowledgeable moral beliefs. But if our high-stakes moral beliefs are to be justified and knowledgeable from the consequentialist point of view, their justification and knowledgeability depends on induction in some way or other. Accordingly, we must inductively project in some way from past experience with a given action type to the conclusion that the target instance of that type will in one's present circumstances maximize net value in the long run.

But the worry immediately arises: it's surely possible that a prospective action that would have horrendous near term consequences never the less maximizes net value in the long run. This is an old objection, and consequentialists have had something to say about this. For example, G. E. Moore (1903 [1988], 93) tentatively suggests the following:

As we proceed further and further from the time at which alternative actions are open to us, the events of which either action would be part cause become

---

because just as prospective actions could have easily failed to maximize net value, they could also have easily failed to satisfy net value (see below).

<sup>6</sup> For example, one needn't be a consequentialist to endorse the following. Suppose it would otherwise be permissible either to help group A or to help group B by providing one with money. However, you know that group B would use a portion of that money to set in place a series of events whose outcomes would very likely severely and unnecessarily harm group A in the long run. Provided helping group A would not have comparably bad likely outcomes, intuitively, helping group B would (*ceteris paribus*) be wrong and it would be wrong because of the likely consequences of doing so. If that intuitive moral judgement is right, it is not one that could be known or justifiably believed if the argument against consequentialism below is sound.

increasingly dependent on those other circumstances, which are the same, whichever action we adopt. The effects of any individual action seem, after a sufficient space of time, to be found only in trifling modifications spread over a very wide area, whereas its immediate effects consist in some prominent modification of a comparatively narrow area. Since, however, most of the things which have any great importance for good or evil are things of this prominent kind, there may be a probability that after a certain time all the effects of any particular action become so nearly indifferent, that any difference between their value and that of the effects of another action, is very unlikely to outweigh an obvious difference in the value of the immediate effects.

J.J. C. (1973, 33) concisely reiterates the idea:

[W]e do not normally in practice need to consider very remote consequences, as these in the end rapidly approximate to zero like the furthest ripples on a pond after a stone has been dropped into it.

Shelly Kagan (1998, 64) says that:

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.

If Moore, Smart, and Kagan are right, then past experience allows us to reliably project the net value of a given action at least in a reasonable range of normal circumstances. For according to their suggestion, the long term effects of any prospective action will (or is objectively likely to) wash out in a way that tends to make its foreseeable near term net value representative of its long term net value. Accordingly, if any kind of action allows for this sort of projection, it's exactly the sort of high-stakes actions described above where the foreseeable near-term consequences are extremely high.<sup>7</sup>

Let's first clarify just how we can move from the informal assertions of projectability made above to outright claims about a given action being right or wrong. Here is an apparently cogent way of specifying the needed details:

CONSEQUENTIALIST MORAL REASONING (CMR):

(1) If up to the present A-ing in circumstances like the ones I'm in have (or would have) frequently enough maximized net value up till now, then A-ing in

---

<sup>7</sup> Yet some have worried about this Moore-Smart-Kagan style response. For it's not hard to imagine a series of events where, say, murdering lots of children has the highest net value in the long run, and it's not quite clear why we should think that the long term consequences will (or are objectively likely to) balance out (Greaves 2016). Moreover, Elgin (2015) points out that, under certain conditions, the mere fact that there are long term distant consequences statistically prohibits us from being able to reliably assess the net value of our prospective actions. But Elgin's criticism does not apply with equal force to all forms of consequentialism. For example, some forms of consequentialism give special place to the interests of beings that already exist (e.g. prior existence utilitarianism, ethical egoism). Once that is done, distant future consequences become irrelevant when it comes to undermining the Moore-Smart-Kagan thesis. An upshot of the epistemic objection I present below is that one cannot retreat to something like prior existence utilitarianism to avoid it.

- my present circumstances will very likely maximize net value in the long run.  
(suggested by Moore, Smart, and Kagan)
- (2) Up to the present A-ing in circumstances like the ones I'm in have (or would have) frequently enough maximized net value up till now.
- (3) Therefore, A-ing in my circumstances *will very likely* maximize net value in the long run. (from 1 and 2)
- (4) Therefore, given that I have no significant reason to think A-ing will fail to maximize net value in the long run, A-ing in my circumstances *will* maximize net value in the long run. (from 3 and contraction (see below))
- (5) An action is wrong iff it fails to maximize net value the long run.  
(consequentialism)
- (6) Therefore, refraining from A-ing in my circumstances is wrong. (from 4 and 5)

Let me say a few things about this pattern of reasoning before getting to the lottery problem for consequentialists.

(1) offers us a way of specifying the underlying idea that Moore, Smart, and Kagan suggested on behalf of consequentialism. The parenthetical "would have" in (1) is to indicate that sometimes we can judge a possible past action's prospective net value in the near-term without the need of anyone having performed that specific action in the past. In no point in the past have 99.9% of earth's population suffered horribly unto death from the release of a virus. Yet we know (or can at any rate be reasonably certain) that if someone were to have done that five years ago, that action would have failed to maximize net value *up to now*—i.e. it's an action whose net value calculated up to the present moment is lower than some alternative prospective action's net value when calculated up to the present moment.

Importantly, the justification for (1) is inductive: provided induction from past experience is sufficiently reliable in the case of high-stakes beliefs, then we have defeasible inductive justification for endorsing the conditional specified by (1). Of course, as noted above, the consequentialist application of (1) assumes that maximization of net value *up to the present* is a sufficiently reliable indicator of maximization of net value *in the long run*. Again, some versions of consequentialism will have a relatively easy time justifying this (prior existence utilitarianism) while other versions will have a comparatively difficult time doing so (classical utilitarianism). This is something that warrants further discussion, but I will pass it by to discuss other issues. For the most part, the justification of (1) becomes progressively easier the more high-stakes our potential actions are irrespective of the version of consequentialism one endorses.

(2) stands to be justified by historical knowledge of the effects of such actions in suitably similar circumstances. (3) is a deductive conclusion from (1) and (2).

(4) relies on what I'm calling a *contraction principle*, i.e. a principle that licenses transitioning from probabilistically qualified claims to probabilistically unqualified claims in the absence of defeating information. For example, when a radiologist examines an X-ray of your leg and concludes that you have a hairline fracture, they are (or can be) implicitly reasoning from a contraction principle. For while X-rays offer us highly reliable representations when interpreted by a trained professional, there is still some small margin of error; there is still some small chance that either the X-ray production involved some error in representation or the reader mistook something for a hair-line fracture that was not a hair-line fracture. Even so, a skilled radiologist can justifiably judge that you in fact have a hair-line fracture, despite the small error possibilities so long as they have no reason to think the small error possibility is actual. Similarly, most of the time a jury judges (and sincerely believes) someone guilty of a crime on the basis of a body of evidence, they engage in a form of contractive reasoning from how things very likely are, to how things actually are. There is a lot to say about contractive reasoning and I will return to this below.

For now, it's enough to note that we regularly engage in (or could engage in) such reasoning, and that often enough such reasoning is justified.

(5) is just a coarse-grained representation of the objective consequentialist thesis. If any of our high-stakes moral beliefs are to be justified on the assumption that some form of consequentialism is correct, then it has to be the case that something akin to a CMR argument underlies the justification of our high-stakes moral beliefs.

### 3 Applied Moral Skepticism

Why think that our high-stakes moral beliefs are like lottery beliefs if consequentialism is true? The answer flows out of the different constraints that Sensitivity and Safety impose on knowledgeable belief.

Take Sensitivity first. Suppose you believe A-ing is wrong on the basis of a CMR argument. Would you still hold this belief on the basis of a CMR argument even if your belief were false, i.e. even if A-ing were not wrong? You would. This is because CMR arguments are not "sensitive" to the falsity of the beliefs they support. The reason for this lies with premise (1), which is a conditional that relies on induction from past experience for its justification. But any reasoning that relies on the past as an indicator of the future will fail to satisfy Sensitivity. For example, believing the sun will rise tomorrow because it has always done so in the past will not count as knowledge since I would still have believed this even if, for whatever reason, the sun were destroyed or the rotation of the earth halted before it had a chance to rise tomorrow (Vogel 1987; Comesaña 2007). Similarly, it could improbably turn out to be the case that murdering all the local school children maximizes net value because that would lead to a distant future where many, many more children are saved from lethal harm they otherwise would have suffered. But even were that the unlikely truth, if my belief that it's wrong were based on a CMR argument I would still have believed that it's wrong to kill all the local school children just on the basis of the fact that such an action would foreseeably fail to maximize net value. Accordingly, if Sensitivity is true then premise (1) is unknown given its inductive justification. And if premise (1) is unknown then presumably we cannot know the moral status of an action on the basis of a CMR argument—at least not those that rely on induction for their justification of premise (1).

Now, it's easy to see how little knowledge remains to us if Sensitivity is true since inductive knowledge becomes exceedingly difficult to come by in general (Vogel 1987; Comesaña 2007). For this reason, many philosophers have been reluctant to endorse Sensitivity and have turned to its contrapositive cousin, Safety, for help. Safety lacks the skeptical implications of Sensitivity for inductive beliefs. So if Safety is true, then premise (1) of CMR is not obviously in jeopardy. And if one can know premise (1) along with the rest of the premises, then it would seem that one could also come by moral knowledge and justified moral beliefs on the basis of CMR arguments also.

The thing to note is that while Safety doesn't threaten premise (1), it does obstruct the derivation of its conclusion by imposing limits on the ability of contractive reasoning to afford us knowledge. For recall that to reach (4) in the CMR we needed to rely on the idea that we can transition from claims about how things *very likely are* (*/will be*) to how things *actually are* (*/will be*). But if Safety is true, then we can only gain knowledge of actuality from knowledge of likelihoods when beliefs so based could not have easily been false. But if consequentialism is true, virtually all of our high-stakes moral beliefs are just that: they could easily have been false.

To see this consider a variation on the Lottery case from above:

#### AGENT-CAUSAL LOTTERY

You're holding a lottery ticket whose number reads: 1524353214. Unlike a standard lottery, the winning ticket number is not determined through a near-random mechanical process. Rather, the winning ticket number is determined by the following process. Each day starting from tomorrow a new participant is selected to decide which number comes next in the series. It can be any



number between 1 and 5. Since there are 10 numbers on the ticket, this will take 10 days and will require 10 participants to select these numbers. Each participant is freely selected by the previous participant, while the first participant is chosen at random. And, excepting the first participant, each participant is informed as to the number selected by the previous participant and encouraged, but not forced, to choose a different number than the previous participant. Accordingly, the participants are influenced by that knowledge when selecting a number.

Given this selection process, it's clear that the winning ticket number will not be chosen at random. So you know that, unlike a fair lottery, the exact chances that your ticket's number will be selected cannot be calculated in any precise way. But you do know that your chance of winning is obviously *very* small. So it's extremely likely that your ticket is a loser. Reflecting on the long odds involved you conclude (L\*) *that your ticket is a loser*. Besides your rough assessment of the odds, you have no other reason to think your ticket is a loser. As it turns out, your number will not be selected in the following ten days and so your belief that you own a losing ticket is true.

This is an "agent-causal lottery" in the sense that there is a clear non-random agent-causal path reaching from the first participant to the winning ticket number. For each participant (except for the first) is selected by the previous participant, and each participant's number selection is causally influenced by their knowledge of the previous participant's number selection (except for the first). Accordingly, the outcome of this lottery is casually produced by the actions performed by previous agents and the responses of the agents who are impacted by those past actions.

Do you know (L\*) *that you have a losing ticket* in Agent-Causal Lottery? I venture to claim that anyone convinced that you cannot know L in Lottery, will be likewise drawn to the conclusion that you cannot know L\* in Agent-Causal Lottery. The parallels are too deep, and the differences are too superficial. But more to the point, if Safety is what we rely on in diagnosing what goes wrong with my belief in Lottery, then we must also rely on it in diagnosing this case. And given the parallels between this case and the original, it's easy to see where this is heading. For in the original case, my belief in L fails to satisfy Safety because it could easily have been false since "all that needs to be different in order for one's ticket to be a winning ticket is that a few numbered balls fall in a slightly different configuration" Pritchard (2008: section 4). And similarly, your belief in L\* in Agent-Causal Lottery could easily have been false since all that needs to be different in order for your ticket to be a winning ticket is for a few people to have made a slightly different decisions about which number and subsequent participant to choose. So if Safety explains my failure to know in Lottery, it explains your failure to know in the Agent-Causal Lottery.

The Agent-Causal Lottery is a trying case for non-skeptical consequentialists. For the outcomes of our high-stakes actions in the long run are a lot like the outcome in an Agent-Causal Lottery. For in both kinds of case there is an agent-causal path that produces the relevant outcome, and at each (or very many) node in that path things could easily have been otherwise. For there are almost always many alternative available prospective actions an agent might easily have performed, thereby leading to different outcomes with a plausibly different net value.

Take, for example, the 2019 mass shooting in El Paso, Texas. 22 were killed and 24 were injured (if this is insufficiently high-stakes just imagine many more were killed and injured). By the lights of ordinary moral intuitions this was a wrong action and we know that this was a wrong action. But the knowledgeability of this as a wrong action doesn't just depend on (i) the known failure of this action to maximize net value in the near term, and (ii) the idea that (i) ensures that the shooting is also *very likely* to fail to maximize net value in the long run. For the lesson of the lottery cases is that beliefs that are highly likely to be true don't necessarily satisfy Safety.

According to Safety, for the belief that the shooting was wrong to constitute knowledge it has to be the case that the shooting could not have easily maximized net value in the long run. But just as in the Agent-Causal Lottery where it's easy to see how one could have easily beat the odds since each person could have easily chosen a slightly different number, in the mass shooting case too one could "easily have beat the odds" in at least some of the following ways. Consider that the El Paso shooting was the most lethal shooting in the US since 1949. And it comes just after a string of other disturbingly violent shootings. It is also a shooting that took place in a highly populated city in Texas—a state that ordinarily shows strong resistance to fire-arm restrictions. Now we know that, historically, social movements that bring about social change often erupt from tragic events. Could it not easily be the case that this shooting helps tip the balance in support of gun-law reform in the US in such a way that future shootings are significantly reduced and thereby maximize net value in the long run? It's hard to see what grounds there could be for resisting this judgement in our present circumstances. But even if that fails to be the case, it could easily be the case that this shooting makes the general population much more vigilant and willing to report on people they know who might be at risk of committing a mass shooting, thus preventing many more shooting, and thereby maximizes net value in the long run. Further, it could easily be the case that this shooting has "identity-affecting" consequences that maximize net value, i.e. it impacts which people come into existence in long run and thereby impacts the net value of one's action. Perhaps, for example, by killing these people in El Paso the shooter has impacted history in a way that would prevent the births of what would otherwise have been the next massively genocidal dictator and thereby maximizes net value in the long run (cf. Lenman 2000).

According to the thesis floated by Moore, Smart, and Kagan, these are very unlikely possibilities since they are possibilities where the near term consequences fail to accurately represent the long term consequences, but so too is my winning the lottery. And what the lottery teaches us is that the unlikelihood of an outcome doesn't ensure that the unlikely outcome could not have easily obtained.

Recall that Sensitivity undermined the knowledgeability of premise (1). This was due to the fact that Sensitivity undermines inductive knowledge generally. But Safety is compatible with inductive knowledge and so it imposes no immediate threat to (1). Rather, Safety obstructs the justification of (4) by imposing a limit on knowledge-generating contractive reasoning. For recall that to reach (4) in the CMR we needed to rely on the idea that we can transition from claims about how things *very likely* are (/will be) to how things *actually are* (/will be). But if Safety is true, then we can only gain knowledge of actuality from knowledge of likelihoods when beliefs so based could not have easily been false. But, given the nature of the relation between our present actions and their outcomes (especially in the social world), consequentialism seems to imply that the vast majority of our high-stakes moral beliefs could easily have been false. This puts knowledge of our high-stakes moral beliefs out of reach.

What impact might this have on the justificatory status of high-stakes moral beliefs if consequentialism is true? Well, if  $J \leftrightarrow K$  or  $J \rightarrow PK$  are true, then our high-stakes moral beliefs are unjustified since both principles limit justified beliefs to those that are potential knowledge. If  $J \rightarrow JPK$  is true, the justificatory status of our high-stakes moral beliefs fare a bit better. Since, so long as one is *ignorant* of the unknowability of our high-stakes moral beliefs it will be easier to have justification to believe that one *is* in a position to know them. This is in some sense "good news" for consequentialists since it carves out space for there to be some unknown yet justified high-stakes moral beliefs. But the consequentialists for whom this is good news are only those who are ignorant of the fact that knowing requires either Sensitivity or Safety.

#### **4 Consequentialism and Self-Defeat**

According to many, the justification for believing (or assigning high credence to) consequentialism or any other general normative theory of ethics depends substantially on its ability to explain our "considered moral judgements", which includes our concrete case moral beliefs (or intuitions) that we reflectively endorse. Our high-stakes moral beliefs form an important subset of our considered

moral judgements about the actual world for they tend to be widely believed and are intuitively striking in the sense that their denials seem clearly to be false. This is doubtless due to the fact that, as defined, our high-stakes moral beliefs involve wantonly harming others in circumstances that have extremely bad effects in the near-term.

Consequentialists have regularly argued that consequentialism can explain the truth of at least an important range of our (correct) considered moral judgements, while offering error theories for those it cannot explain. Let ‘HSMBs’ refer to our high-stakes moral beliefs about the moral status of prospective actions *that consequentialism can explain*. This will include only those HSMBs that are true by the lights of consequentialism. For example, if the HSMB is *that it’s wrong to murder Sam*, it is an HSMB that consequentialism can explain only if killing Sam fails to maximize net value in the long run. Otherwise, it’s just not true and so not the sort of HSMB that a consequentialist should expect their theory to explain. Explaining why people have *false* HSMBs is the job of an error theory.

As is common, I interpret the evidential relevance of explanatory considerations in probabilistic terms. That is, we are to understand data that is explained by a hypothesis as increasing the likelihood of that hypothesis. Thus:

$$\Pr(\text{Consequentialism} \mid \text{HSMBs}) > \Pr(\text{Consequentialism})$$

But this evidential inequality is only one part of the story of how we might be able to come to justifiedly *increase* our confidence in consequentialism in light of our HSMBs. For in order to justifiedly increase our confidence in consequentialism we must be able *to update* (e.g. conditionalize, jeffrey conditionalize, pseudo-conditionalize) on our HSMBs. But on all accounts, in order to update on some evidence, E, we have to stand in some *epistemically significant* relation to E. For example, standard update rules require that we have *learned* E. Now, it’s an interesting question what it takes to “learn” that E. But at a minimum it should require that one have *justification* for believing E. The idea that one can justifiedly update on information that one doesn’t even have justification to believe is very hard to make sense of.

Now, the arguments of the previous section show that knowledge-centric anti-luck epistemology is inconsistent with having justification for believing our HSMBs, and so knowledge-centric anti-luck epistemology is inconsistent with the justification of updating on our HSMBs *if consequentialism is true*. This threatens to yield a form of standard first-order epistemic defeat for consequentialism since a crucial part of the evidence that is supposed to justify it, our HSMBs, are inaccessible to us.

Yet consequentialism might be false. If consequentialism is false, then our HSMBs can be justified provided the correct normative theory of ethics doesn’t give the long term consequences of our actions a role in determining right from wrong (and thereby run afoul of knowledge centric anti-luck epistemology). Generally, non-consequentialist theories of ethics and rule consequentialist theories don’t do this, and thereby create a more hospitable environment for the justification of our HSMBs. Now, if our HSMBs can be justified then presumably they can also be justifiedly updated on, and therefore they *can* function as evidence for consequentialism! This is a surprising little fact, one that gives us a bit of *higher-order information about our evidence for consequentialism*, namely:

HOE: Our evidence (constituted by our HSMBs) supports having an increased degree of confidence in consequentialism only if consequentialism is false.

Now on the probabilistic outlook I began with the question of having a justified high credence in consequentialism depends on the evidential relation between our prior confidence in consequentialism and our prior confidence in consequentialism conditional on our HSMBs.

Having observed the way in which HOE follows from the arguments above, our evidence now includes both our *HSMBs and HOE*.

Intuitively, HOE should have some evidential impact on our confidence in consequentialism. But what kind of impact, exactly? Surely it should not *increase* our confidence in consequentialism. Put in general terms, it's beyond credulity to think one can know/justifiedly believe that (i) E, and that (ii) E supports P only if  $\neg P$ , and think that one can justifiedly believe (or increase confidence in) P on the basis of (i) and (ii).<sup>8</sup> That leaves two options with regard to the evidential impact of HOE: either HOE *lowers* our posterior confidence in consequentialism, or it *screens off* the relevance of our HSMBs. I don't know of an uncontroversial reason to prefer either disjunct. So let us remain neutral on this issue. Accordingly, we have the following inequality:

$$\Pr(\text{Consequentialism} \mid \text{HSMBs \& HOE}) \leq \Pr(\text{Consequentialism})$$

That is to say, learning HOE at the very best screens off whatever justification our HSMBs afforded us for thinking that consequentialism is true; at worst it should lower our credence in consequentialism.

So unless one has sufficient reason to believe (or at least increase confidence in) consequentialism that is wholly independent of our HSMBs, consequentialism is not a moral theory that we can justifiedly believe or have high confidence in. Put differently, *if* justifiedly believing consequentialism depends on its ability to explain our HSMBs, then consequentialism doesn't seem like the sort of ethical theory that can be justifiedly believed. It's a kind of blindspot, a truth about the structure of moral normativity that we would seem incapable of rationally recognizing as such.

There are a number of questions this raises about the methodology of justifying normative theories. I think the most salient one it raises concerns the ability of concrete case judgments or intuitions about merely possible cases to justify normative theories. For when it comes to the merely possible cases we can generally specify in the very construction of the cases whether or not the prospective action under consideration maximizes net value in the long run. If ethical theorizing can function in an epistemically robust way with only such cases to work with, then the applied moral skepticism of consequentialism would not impose a limit on our ability to justifiedly increase our confidence in consequentialism. But, as others have worried, there seems to be something epistemically circular about this.<sup>9</sup> Whether or not it is an epistemically problematic form of circularity is a discussion for another time.

It's worth mentioning that there is one form of consequentialism that may evade these worries: prior existence consequentialism. For on such views whether or not an action is right or wrong just depends on its impact on individuals who already exist (or would exist no matter which action were performed). Now, unlike our prospective actions, we have historical knowledge of the actual outcomes of people's past actions and how those actions impacted the people who existed (or doubtless would exist) at that time. Arguably, this historical knowledge can afford us knowledge of the moral status of past high-stakes actions if a version of prior existence consequentialism is correct. For example, the dropping of a second atomic bomb on a populated area like Nagasaki was gratuitous for quickly ending war with Japan. The war could have been brought to just as quick an end if it were not dropped or if it were dropped on an uninhabited area of Japan. We know this. So we know that this past action was wrong, even if at the time those who were making the decision were not—by the lights of knowledge-centric anti-luck epistemology—able to know or

---

<sup>8</sup> If the evidential support relation involved some kind of standard conditional (material, indicative, counterfactual, strict), then (i) and (ii) would yield a contradiction. For whichever way the conditional is interpreted, it cannot be self-consistent and non-trivially true that:  $(E \& (E \rightarrow P)) \rightarrow \neg P$ . But I'm here assuming evidential support is not essentially bound to such conditionals, and can often be understood in broadly probabilistic terms.

<sup>9</sup> See Pust's (2013, Ch. 1) discussion of Rawls and the method of reflective equilibrium.

justifiedly believe it. The upshot is that prior existence consequentialist views may only imply a limited form of skepticism: we may not be able to know or justifiedly believe whether or not a prospective action of *ours* is wrong (due to the arguments of section 3), but at least we can justifiedly increase our confidence in prior existence consequentialist views in virtue of their explanatory power with respect to HSMBs about the past.<sup>1011</sup>

## 5 References

- Bird, Alexander. 2007. "Justified Judging." *Philosophy and Phenomenological Research* 74 (1): 81–110. <https://doi.org/10.1111/j.1933-1592.2007.00004.x>.
- Comesaña, Juan. 2007. "Knowledge and Subjunctive Conditionals." *Philosophy Compass* 2 (6): 781–91. <https://doi.org/10.1111/j.1747-9991.2007.00076.x>.
- Elgin, Samuel. 2015. "The Unreliability of Foreseeable Consequences: A Return to the Epistemic Objection." *Ethical Theory and Moral Practice* 18 (4): 759–66. <https://doi.org/10.1007/s10677-015-9602-8>.
- Greaves, Hilary. 2016. "Cluelessness." *Proceedings of the Aristotelian Society* 116 (3): 311–39. <https://doi.org/10.1093/arisoc/aow018>.
- Ichikawa, Jonathan Jenkins. 2014. "Justification Is Potential Knowledge." *Canadian Journal of Philosophy* 44 (2): 184–206. <https://doi.org/10.1080/00455091.2014.923240>.
- Kagan, Shelly. 1998. *Normative Ethics*. Boulder, CO: Westview Press.
- Kelp, Christoph. 2018. *Good Thinking: A Knowledge First Virtue Epistemology*. New York, NY: Routledge.
- Lenman, James. 2000. "Consequentialism and Cluelessness." *Philosophy & Public Affairs* 29 (4): 342–70.
- Littlejohn, Clayton. 2012. *Justification and the Truth-Connection*. Cambridge: Cambridge University Press.
- Millar, Alan. 2019. *Knowing by Perceiving*. Oxford: Oxford University Press.
- Miracchi, Lisa. 2015. "Competence to Know." *Philosophical Studies* 172 (1): 29–56.
- Moore, George Edward. 1903 [1988]. *Principia Ethica*. Amherst, NY: Prometheus Books.
- Pritchard, Duncan. 2005. *Epistemic Luck*. Oxford: Oxford University Press.
- Pritchard, Duncan. 2008. "Knowledge, Luck, and Lotteries." In *New Waves in Epistemology*, edited by Vincent F. Hendricks and Duncan Pritchard, 28–51. Basingstoke: Palgrave Macmillan.
- Pritchard, Duncan. 2012. "Anti-Luck Virtue Epistemology." *The Journal of Philosophy* 109 (3): 247–79. <https://doi.org/10.5840/jphil201210939>.
- Pust, Joel. 2013. *Intuitions as Evidence*. New York, NY: Routledge.
- Rabinowitz, Dani. 2019. "The Safety Condition on Knowledge." <https://www.iep.utm.edu/safety-c/>.
- Rosenkranz, Sven. 2018. "The Structure of Justification." *Mind* 127 (506): 309–38. <https://doi.org/10.1093/mind/fzx039>.
- Silva, Paul. 2017. "Knowing How to Put Knowledge First in the Theory of Justification." *Episteme* 14 (4): 393–412. <https://doi.org/10.1017/epi.2016.10>.
- Singer, Peter. 2011. *Practical Ethics*. Cambridge: Cambridge University Press.
- Smart, John Jamieson Carswell, and Bernard Williams. 1973. *Utilitarianism: For and Against*. Cambridge: Cambridge University Press.
- Smithies, Declan. 2012. "The Normative Role of Knowledge." *Noûs* 46 (2): 265–88. <https://doi.org/10.1111/j.1468-0068.2010.00787.x>.

---

<sup>10</sup> For obvious reasons, this historical knowledge is not helpful for versions of consequentialism, like classical utilitarianism, which make the moral status of past actions depend on distant future outcomes. I'm grateful to Rik Peels for prompting me to reflect on whether our historical knowledge of outcomes could be useful in the justification of consequentialist views.

<sup>11</sup> I'm grateful to the Alexander von Humboldt Foundation which funded this research as well as to Michael Blome-Tillmann, Michael Klenk, and Rik Peels who offered helpful feedback on this project.

- Sutton, Jonathan. 2007. *Without Justification*. Bradford Books. Cambridge, MA: MIT Press.
- Vogel, J. 1987. "Tracking, Closure, and Inductive Knowledge." In *the Possibility of Knowledge: Nozick and His Critics*, edited by Steven Luper, 197–215. Lanham, MD: Rowman & Littlefield.
- Williamson, Timothy. 2013. "Knowledge First." In *Contemporary Debates in Epistemology*, edited by Matthias Steup, John Turri, and Ernest Sosa, 1–9. Hoboken: Wiley-Blackwell.