

Forthcoming in *American Philosophical Quarterly*

## Is Reliabilism a Form of Consequentialism?

Jeffrey Dunn

Kristoffer Ahlstrom-Vij

**Abstract:** Reliabilism—the view that a belief is justified iff it is produced by a reliable process—is often characterized as a form of consequentialism. Recently, critics of reliabilism have suggested that, since a form of consequentialism, reliabilism condones a variety of problematic trade-offs, involving cases where someone forms an epistemically deficient belief now that will lead her to more epistemic value later. In the present paper, we argue that the relevant argument against reliabilism fails because it equivocates. While there is a sense in which reliabilism is a kind of consequentialism, it is not of a kind on which we should expect problematic trade-offs.

### 1. Introduction

Reliabilism about justification maintains that a belief is justified iff it is produced by a reliable belief-forming process. Though there are differences of detail, the basic idea is that a reliable process is one that generates a high ratio of true to false beliefs.<sup>1</sup> Thus, reliabilism maintains that processes capable of producing justified beliefs are those that tend to produce true beliefs. For this reason, it has often been thought to be a kind of epistemic consequentialism: true belief has epistemic value, and the processes that generate justification are those that tend to get the believer this value.<sup>2</sup>

Recently, it has been suggested that reliabilism should be rejected because it has us condone a variety of problematic trade-offs (Berker 2013a, b and Littlejohn 2012; see also Jenkins 2007 and Firth 1981). The relevant trade-offs involve cases where someone forms what is intuitively an epistemically deficient belief that nevertheless leads her to more epistemic value than if she hadn't formed the deficient belief. For instance, suppose that, if I form what I have good reason to believe is a false belief now, five of my other beliefs will definitely be true. If reliabilism maintains that what matters for justification is the production of true belief, it looks like it might condone such trade-offs. But such trade-offs often seem as though they shouldn't be condoned: it seems you epistemically shouldn't form a belief you have good reason to believe is false even if the result of doing so is more true belief.

Those presenting the relevant trade-off problems take themselves to be exploiting a defining feature of reliabilism: its commitment to *consequentialism*. The general argument is as follows:

**Trade-Off Argument (TOA)**

- (1) Consequentialists condone problematic trade-offs.
- (2) Reliabilism is a form of consequentialism.
- (3) Hence, reliabilists condone problematic trade-offs.

TOA is valid, and (1) and (2) seem *prima facie* plausible: as we shall see below, consequentialists in ethics have been shown to condone problematic trade-offs (section 2), and reliabilism has repeatedly been described as a form of consequentialism by defenders and opponents alike (section 3). Hence, her critics infer (3) and maintain that the reliabilist will condone problematic trade-offs between *propositions*, just like her ethical cousin will condone problematic trade-offs between *persons*. Here, for example, is Berker:

[...] we are working with the hypothesis that the reason why many epistemic consequentialists are susceptible to Firth-style counter-examples [in terms of problematic trade-offs] is precisely because they countenance cross-propositional trade-offs in their determination of the epistemic status of individual beliefs (2013b, p. 372).

Littlejohn (2012, p. 82) makes a similar claim against what he calls “Epistemic Rule Consequentialism” of which reliabilism about justification is an exemplar. On Littlejohn’s way of running the argument, it is reliabilism’s commitment to what he terms the totalizing assumption, “the assumption that total intrinsic value determines deontic status” (p. 80), that generates problems by having the reliabilist, and indeed all epistemic consequentialists, condone objectionable trade-offs.

Reliabilists have in turn argued that (3) is false (Ahlstrom-Vij and Dunn 2014 and Goldman 2015). Strictly speaking, however, the relevant defenses show no such thing; all they show is that the trade-off cases *presented to date* do not have to be condoned by the reliabilist, which falls short of showing that (3) is false. More to the point, since TOA is valid and (1) and (2) look plausible, we should expect (3) to be true—and if so, the reliabilist has reason to worry that there are bound to be forthcoming cases that might not be as easily dealt with. This is the worry we will be concerned with in the below.

So how worried should reliabilists be? Not particularly, if we are right: TOA equivocates. Any plausible understanding of consequentialism that makes (1) true, makes (2) false, and vice versa. If we are right about this, there is no reason to believe that (3) holds—nor should we expect that further work on the part of the anti-reliabilist will reveal any problematic trade-offs for reliabilism.

## 2. Trade-off Problems in Ethics

Why think that premise (1) in TOA is true? For one thing, consequentialists have been shown to condone problematic trade-offs in ethics. Consider the infamous surgeon case: the classic utilitarian seems to sanction cutting up a healthy person—even if against her will—if that saves the lives of five people in need of organs (Thomson 1976, p. 206). Still, most of us find that clearly immoral. We have the same reaction in other well-known cases in the literature, such as that involving the framing of a perfectly innocent person to prevent a mob from starting a violent riot (McCloskey 1957, pp. 468-469), or a person being offered to shoot (Williams 1973, pp. 98-99) or torture (Foot 1985, pp. 197-198) one person to save a large number of people.

Crucially, according to her critics, the utilitarian's condoning of the relevant trade-offs is not a mere *accident* of her theory. As Foot (1985, p. 196) puts the point: "what is most radically wrong with utilitarianism is its consequentialism." The consequentialist, the story goes, fails to respect "the separateness of persons," as Rawls (1963, p. 124) puts it, and ends up sanctioning the promotion of utility even when it violates people's rights. Borrowing a term from Nozick (1974, 1981), we can formulate the relevant complaint in terms of the consequentialist's blindness to *side-constraints* on actions. To say that there are side-constraints on actions is to say that it's sometimes right *not* to do what has better (including the best) consequences. Williams gets to the same idea in his critique of utilitarianism:

All that is involved on the present account, in the denial of consequentialism, is that with respect to some type of action, there are some situations in which that would be the right thing to do, even though the state of affairs produced by one's doing that would be worse than some other state of affairs accessible to one. (Williams 1973, p. 90)

Foot (1985) gives voice to a similar idea when suggesting that what is compelling about consequentialism—and what also makes it so problematic—is “the rather simple thought that it can never be right to prefer a worse state of affairs to a better” (p. 198).

Now, we are not suggesting that it’s impossible to understand “consequentialism” in a manner different from the way aforementioned critics do. Our purpose here is simply to try to understand what type of consequentialism philosophers worrying about problematic trade-offs are concerned with, and to thereby get clearer on the notion of consequentialism at work in (1). As we have seen, the type of consequentialism figuring in these discussions is one that denies the existence of side-constraints. Hence, we get the following criterion:

**Side-Constraints:** A theory of rightness is consequentialist iff it imposes no side-constraints.

Side-Constraints seems a plausible criterion for two related reasons. First, it makes clear the distinction between consequentialist theories and theories in the Kantian tradition, with the former rejecting and the latter accepting side-constraints. As Nozick (1981, pp. 30-31) writes, “[s]ide constraints upon action reflect the underlying Kantian principle that individuals are ends and not merely means; they may not be sacrificed or used for the achieving of other ends without their consent.” Second, as we have seen, in understanding consequentialism as rejecting side-constraints, it goes exactly to what critics have found problematic about consequentialist theories in ethics: if we reject side-constraints, we have to condone actions that seem clearly wrong, precisely because those actions have very good consequences. That is, we have to accept problematic trade-offs.

In the next section, we will therefore consider whether reliabilism is consequentialist in the sense captured by Side-Constraints. Our answer will be *no*. There are, however, other no-

tions of consequentialism, and so after this we will consider whether any of these notions save TOA—our answer to that, too, will be *no*.

### 3. Reliabilism and Side-Constraints

Premise (2) in TOA states that reliabilism is a form of consequentialism. This is certainly in line with what many epistemologists have suggested. For example, Firth (1981, p. 12) writes that “reliability theories of justification [...] might well be described as forms of epistemological rule-utilitarianism.” More recently, Berker (2013b, p. 350) writes that, “[j]ust as utilitarianism is the paradigm example of a teleological ethical theory, so too, I believe, reliabilism is the paradigm example of a teleological epistemic theory.” Along similar lines, Percival (2002) suggests that “[e]pistemic consequentialism is the epistemic analogue of ethical consequentialism” (p. 121), and that reliabilism “can be profitably subsumed under” the former (p. 125). Chase (2004, p. 124) characterizes Goldman’s reliabilism as “an epistemological form of rule consequentialism,” which moreover is in line with what Goldman (1986, p. 97) himself says on the matter.

That said, if what we said in section 2 is correct, we need to dig deeper than this in order to determine whether (2) is true. More specifically, in order not to run the risk of equivocating, we need to ask: Is reliabilism a form of consequentialism *in the sense of “consequentialism” at work in (1)*? That is, does reliabilism satisfy Side-Constraints?

As we saw in the previous section, in the ethical literature, the idea of side-constraints is tied to the rights of persons: there are certain things you cannot do to people without violating their rights. That is what gives rise to side-constraints and makes some inter-personal trade-offs objectionable. Since neither propositions nor beliefs have rights, this line of thought doesn’t translate directly to the epistemic domain. However, on a more abstract level, we can make sense of an epistemic analogue of the idea behind Side-Constraints: *there are some*

*things you are not to believe, even if doing so would have better consequences—including extremely good consequences—from the perspective of epistemic value.* But the reliabilist agrees with that. Consider the following example: Assume that your coming to believe some proposition  $p$  by guessing would have as a causal consequence your formation of a large number of true beliefs that you otherwise would not have formed. Reliabilism implies that you would *not* be justified in believing  $p$  in this manner, since guessing is an unreliable way of forming belief. This is so despite the fact that believing  $p$  would lead to what reliabilism claims is of final epistemic value (i.e., true belief). Consequently, the reliabilist takes there to be side-constraints: she believes that there are certain things you are not to believe, even if doing so would have very good epistemic consequences.

It might be objected that the reason reliabilism doesn't condone guessing is that, even in the type of case imagined, guessing leads to a lot of false beliefs too.<sup>3</sup> But that misses our point. Reliabilism simply doesn't care about the consequences of *beliefs*, whether (epistemically) good or bad. That is, reliabilism is not an epistemic analogue of act utilitarianism. Reliabilism only cares about the tendency of the *process* to generate true or false beliefs. Differently put, reliabilism looks *backwards*, not *forwards*, which is why Goldman (1979) refers to the relevant type of process reliabilism as "historical reliabilism". Once we see that, we also see that reliabilists think there are some things you shouldn't believe, even if so doing would have very good consequences, the reason being that they simply don't factor in the consequences of forming beliefs when determining whether a belief is justified.

There are, of course, many other types of trade-offs that we can imagine putting to the reliabilist—which is exactly what her opponents have been doing (e.g., in Firth 1981, Littlejohn 2012, and Berker 2013*a, b*). However, the fact that none of those trade-offs are, in fact, ones sanctioned by the reliabilist (see Ahlstrom-Vij and Dunn 2014 and Goldman 2015) simply reinforces the idea that reliabilists *do* endorse epistemic side-constraints: according to

the reliabilist, there are things you should not believe, even if so doing would have better—including very good—consequences. For a comprehensive defense of that claim, the reader may consult the above references, but to get a sense of the type of arguments involved, we may consider two types of cases:

First, consider a case where it looks like the reliabilist would have to sanction a problematic trade-off, but in fact does not. For example, consider Berker's (2013*b*) John Doe case. As Berker sets it up, John Doe's belief that he will recover from a fatal illness—a belief held in the face of his doctors telling him otherwise—will have as a causal consequence that he lives long enough to form a large number of true beliefs. We might think that the reliabilist, being an epistemic consequentialist, would have to say that Doe's belief that he will recover is justified, thereby trading the intuitive “bad” of believing against expert advice for the “good” of forming lots of true belief. But that's not so. Again, reliabilism doesn't evaluate the epistemic status of a belief with reference to its consequences, but with reference to the process that generates it. Arguably, stubbornly clinging to a belief about one's health in the face of contrary testimony from medical professionals is not a reliable way to form, or to maintain, beliefs. Hence, the belief is not justified.

Second, consider a case where the reliabilist sanctions a trade-off that might at first look problematic, but that under closer scrutiny is not. Specifically, consider Berker's prime number case:

Suppose the following is true of me: whenever I contemplate whether a given natural number is prime, I form a belief that it is not. “Is 25 prime? No, it is not.” “Is 604 prime? No, it is not.” “Is 7 prime? No, it is not.” Let us also stipulate that this is the only cognitive process by which I form beliefs about the primeness of natural numbers [. . .]. Since the ratio of prime to composite numbers less than  $n$  approaches 0 as  $n$  approaches infinity



ty, my belief-forming process tends to yield a ratio of true to false beliefs that approaches 1. Therefore process reliabilists are forced to say that, because my belief-forming process is almost perfectly reliable, any belief formed on its basis is justified. But that's crazy! When I form a belief that 7 is not prime, it is simply not correct to say that, although that belief is false, it is epistemically redeemed by the truth of the other beliefs which would be formed via the process that led to it (Berker 2013*b*: pp. 374 and 375).

The first thing to note is that Berker hasn't defined the process in question. All that we have been told is that there's a process such that it's the only one he uses when forming beliefs about primes. But many different processes could have this feature, including one that issues verdicts about all sorts of mathematical properties of natural numbers and behaves the way Berker has described only in the situation in which the query is whether a number is prime. Nothing about the density of primes establishes that *this* process is reliable, since the scope of its application is far wider than the domain of primeness. So, presumably, what Berker wants is to define a process that is *dedicated* to forming beliefs about primes. Only then does the density of non-primes in the interval from 0 to infinity tell us anything about the reliability of the relevant process. In other words, Berker wants us to assume that (a) there are processes dedicated to generating beliefs about primeness, (b) he is relying on such a dedicated process—let us refer to it as *P*—in the scenario imagined, (c) for any number queried, *P* generates the output that it is not prime, and (d) numbers are queried in some quasi-random way among the natural numbers. Under those assumptions, *P* is reliable. And if so, reliabilism should say that the belief *that 7 is not prime*, generated by way of *P*, is justified. That, we claim, is the correct verdict. Any feeling that this is a counter-intuitive verdict should be traced, not to reliabilism, but to the psychological implausibility of (a), especially when

paired with (d). For note that it is implausible indeed that some agent is as likely to contemplate whether 73,046,482,192,753 is prime as whether 53 is prime.<sup>4</sup>

To summarize, reliabilism is definitely *not* consequentialist according to Side-Constraints: reliabilists believe that there are things you should not believe, even if so doing would have better—including very good—consequences. Now, it does not follow that because a view is not consequentialist according to Side-Constraints it sanctions *no* trade-offs. But, as we've suggested, there is reason to believe that the alleged *problematic* trade-offs for reliabilism are either not trade-offs or not problematic.<sup>5</sup> For that reason, it should come as no surprise that the trade-off cases put to the reliabilist so far present no problem for her. Consequently, if we understand consequentialism according to the Side-Constraints criterion, premise (1) in TOA is true but premise (2) is false, which means that we have no reason to believe that (3) holds.<sup>6</sup>

#### 4. Reliabilism and Dependency

Is there an alternative way to understand consequentialism that has both premises of TOA come out true? Let us consider another influential criterion for consequentialism, on which the way to mark the distinction between consequentialism and non-consequentialism is with reference to certain *reductive* claims. Michael Smith writes:

Consequentialism's core claim thus turns out to be about the metaphysics of normativity: normative facts of one sort, facts about the deontic status of actions, are reducible to normative facts of a quite different sort, facts about the values of the possible worlds in which those actions and their alternatives are performed (Smith 2009, p. 258).

The idea here is that consequentialism takes facts about rightness to reduce to facts about the goodness of possible worlds. This is a strong claim. If facts about *X* reduce to facts about *Y*, this often is taken to mean that we have shown that there really are no *X*s; they are *eliminated* in the process of reduction. But consequentialists about rightness need not be committed to the claim that there is no property of rightness. We could slightly weaken the view, however, and hold that consequentialists are committed, not to a reductive thesis, but rather to a *dependency* thesis. This is the view we find in Parfit (2011, p. 373), who, being concerned with the rightness of acts in particular, suggests that consequentialists hold that “[w]hether our acts are right or wrong depends only on facts about how it would be best for things to go.” Offering a slightly more general characterization along similar lines, Sinnott-Armstrong (2011) writes that “[c]onsequentialism [...] is the view that normative properties depend only on consequences.” This gives us the following criterion:

**Dependency:** A theory of rightness is consequentialist iff it maintains that facts about rightness depend solely on the goodness of consequences, be it of individual actions, of accepting or complying with certain rules, or the like.

Two things about this. First, the relevant notion of dependency is meant to be general enough to cover both reductive and non-reductive dependency relations (such as supervenience). Second, Dependency talks about rightness as depending *solely* on the goodness of consequences, and as such is not committing the consequentialist to the idea that rightness needs to depend on *all* consequences. For example, act utilitarians define rightness in terms of the consequences *of performing the act evaluated*. Rule utilitarians instead define right in terms of the consequences *of people accepting or complying with a certain rule*. Still, so long as the relevant positions define rightness solely in terms of the good of consequences—be it

of actions, rules, or what have you—they can helpfully be referred to as ‘consequentialist’, on this criterion.

Understood in this manner, Dependency is a plausible criterion. It gets to the familiar idea that the consequentialist grounds the “right” in the “good” and so emphasizes the contrast traditionally seen between consequentialism and its main, non-consequentialist rivals. Deontological theories, of the sort we find in Kant, reverse the order by grounding the “good” (and the “good will” in particular) in the “right” (as captured by universalisable maxims), while virtue ethical theories ground the “right”, not in the “good”, but in the type of conduct that the virtuous person would engage in.

How does Dependency apply to reliabilism? Fairly straightforwardly: the reliabilist maintains that facts about the epistemic rightness (i.e., justification) of a belief depend solely on the goodness of the consequences of the belief-forming process that led to that belief. So, reliabilism comes out consequentialist on the Dependency criterion. If that notion were invoked in our reading of (2), the premise would come out true. But what about premise (1)? On this reading of “consequentialism”, is it true that consequentialism must condone problematic trade-offs?

It’s not clear that it is. Granted, we can certainly imagine ways of spelling out a consequentialist theory in epistemology that would both satisfy Dependency and condone problematic trade-offs. Consider the following epistemic analogue of act utilitarianism, for example: a belief is justified iff the consequences of forming that belief maximizes epistemic value. This theory would satisfy Dependency and condone problematic trade-offs, for example in cases where forming some belief in a paradigmatically irresponsible manner now maximizes epistemic value in the future. But this doesn’t go to show that a position satisfying Dependency *entails* that it condones problematic trade-offs. For such entailment to hold, it

would need to be the case that *every* position satisfying Dependency condones problematic trade-offs. Do we have any reason to believe that?

No. Consider the following account of justification: on Immediate Consequence Maximization (ICM), a belief is justified iff the *immediate* consequences of forming that belief maximizes epistemic value where the *immediate* consequences of forming a belief consist of just the belief itself. ICM is probably not a very good account of justification (for one thing, it collapses the distinction between justification and truth), but the point of introducing it here is not to defend it; it's simply to illustrate the fact that a view can satisfy Dependency while not accepting any problematic trade-offs. After all, ICM will not condone *any* trade-offs—let alone any *problematic* trade-offs—between losses and gains, for the simple reason that it doesn't factor in anything beyond the epistemic value of the belief itself when determining that belief's degree of justification. And if that's so, we see that a view satisfying Dependency *need not* condone problematic trade-offs.

It might be objected that we are here ignoring trade-offs that are not temporal in nature. What about, say, trade-offs between the belief formed and other beliefs already held by the agent?<sup>7</sup> As it happens, by only factoring in the belief itself, ICM does not take into account the effects on gains and losses with respect to beliefs you already hold, just like it doesn't factor in gains and losses with respect to future beliefs. That's because it only considers the epistemic value of the belief itself. You might think that makes for a bad theory of justification, of course, and we're not denying that; again, we are simply looking to demonstrate that satisfying Dependency doesn't entail accepting trade-offs.

At this point, it might be objected that ICM does not actually satisfy Dependency, since a belief is not a consequence of itself. But this objection is misguided. To see why, consider what we may call the *consequence set* of a belief *B*, consisting in all doxastic consequences of *B*. In this set might surely be other beliefs that the belief that *B* has led to. When looking to

see the ratio of true to false beliefs in this set, we will look to see whether these beliefs are true or false. But surely, we also will want to include *B* itself and whether it is true or false. And if so, while *B* is not a consequence of itself, *B* is in *B*'s consequence set. Furthermore, there's nothing *non-consequentialist* about restricting a consequence set for purposes of consequentialist evaluation. For example, a consequentialist might include consequences until the end of time, or until some time short of that. What decision is made here will impact how *plausible* the view will come out, but it won't impact whether or not it's a *consequentialist* view to start with. As a result, since *B* is in its own consequence set, and there's nothing non-consequentialist about restricting the size of that set—including, in the limiting case, restricting it all the way down to *B*—ICM does count as consequentialist, at least according to Dependency.<sup>8</sup>

Naturally, the fact that a view can satisfy Dependency and not condone problematic trade-offs doesn't go to show that *reliabilism* doesn't condone any problematic trade-offs. But it does mean that citing reliabilism's commitment to consequentialism, in the sense given by Dependency, is not enough to establish that it does condone such trade-offs. Further, it is worth asking at this point on whose shoulders the burden of proof lies. Had reliabilism satisfied Side-Constraints, and as such denied that there were any side-constraints, we would have had reason to expect the reliabilist to condone problematic trade-offs, given the anti-consequentialist's claims that it's exactly that denial that generates the relevant problems. But reliabilism doesn't satisfy Side-Constraints. Had a position satisfying Dependency entailed that it condones problematic trade-offs, then we would have had reason to expect the reliabilist to condone such trade-offs, since reliabilism satisfies Dependency. But that entailment doesn't hold. Consequently, unless the anti-reliabilist is able to come up with some reason to believe that positions should be expected to condone problematic trade-offs *simply on ac-*

*count of satisfying Dependency*, we conclude that, on a Dependency reading of “consequentialism”, (2) comes out true but (1) false.

### 5. Reliabilism, Agent-Neutrality, Maximization and Optimality

Someone who still believes that there’s a way to read TOA that has both (1) and (2) come out true without equivocating on “consequentialism” might want a more comprehensive survey of possible notions of consequentialism. Consider, then, the suggestion that consequentialist theories are purely *agent-neutral*, as opposed to *agent-relative* (McNaughton and Rawling 1991 and Howard-Snyder 1994). To get a sense of these terms, consider the operator “*should ensure the state of affairs [...] obtains,*” where the ellipsis is to be filled in with a description of a state of affairs. If the agent involved makes an ineliminable occurrence in the description of the state of affairs, then the injunction is agent-relative; otherwise it is agent-neutral. For instance, consider the following examples:

*Agent-neutral:* Jim should ensure the state of affairs *where lying is minimized* obtains.

*Agent-relative:* Jim should ensure the state of affairs *where Jim does not lie* obtains.

The first (agent-neutral) injunction allows that Jim could himself lie if it will minimize overall lying. The second injunction forbids this. On this view of consequentialism, consequentialist views issue no agent-relative injunctions except those that might be entailed by agent-neutral ones.

However, this criterion is implausible as a characterization of consequentialism. We can imagine a view that says that it is permissible for agents to act in whatever way maximizes their own utility. That would be a kind of utilitarian egoism. On Agent-Neutrality, this theory would come out non-consequentialist. But that seems wrong. It might be a *bad* theory (we

can think of many instances where an agent is *not* permitted to maximize her own utility), but why think it's not a *consequentialist* theory? It takes rightness to depend solely on consequences, and (we can imagine) does not impose any side-constraints. So why think that it is not a consequentialist theory? We see no reason to think that it is not, which is why we reject Agent-Neutrality as a plausible criterion.

Another way to distinguish between consequentialism and non-consequentialism focuses on *maximization*. For instance, here is Brown:

[W]hatever is meant by 'consequentialism', it must be intelligible as an elaboration of the familiar consequentialist slogan "Maximize the good." The non-negotiable core of consequentialism, I shall assume, is the claim that an action is right, or permissible, iff it maximizes the good. (Brown 2011, p. 751)

This criterion seems less than fully plausible, too. Granted, it fits well with Mill's (1861/1987) utilitarianism, which ties moral obligation (and thereby also permissibility) to the maximization of utility. But Maximization disqualifies satisficing utilitarianism (Slote 1984) as a form of consequentialism. Satisficing utilitarianism takes rightness to depend solely on consequences, and doesn't (or at least doesn't *have* to) impose any side-constraints. The same goes for Bentham's (1789/1961) utilitarianism, which understands the permissibility of an action in terms of how "the tendency it has to augment the happiness of the community is greater than any it has to diminish it" (p. 66). That merely entails that a permissible action generates a net gain of happiness, not that it maximizes it. So why should the mere fact that neither the satisficing utilitarian nor Bentham ties permissibility to maximization disqualify either as a form of consequentialism? For no good reason, as far as we can see.



At the same time, something in the neighborhood of what Brown is appealing to is a plausible criterion for consequentialism. Consider Portmore, who thinks of the consequentialist as being committed to the following idea:

It is always permissible (in the knowledge-supposing sense) for an agent to act so as to bring about the highest ranked available outcome, i.e., the outcome that she has better object-given reasons to prefer above all other available alternatives (Portmore 2007, p. 50).<sup>9</sup>

According to Brown, the consequentialist must say that an action is permissible if *and only if* it maximizes the good. Portmore, on the other hand, requires that the consequentialist say something strictly weaker: that maximization of the good is (merely) *sufficient* for permissibility. This yields the following criterion:

**Optimality:** A theory of rightness is consequentialist iff it says that it is always permissible to maximize the good.

Optimality seems more plausible as a criterion than does Maximization, if only because it counts not only Mill's utilitarianism but also Bentham's and the satisficing utilitarian's theories as consequentialist. That's a good reason to reject Maximization in favor of Optimality in our view. But does reliabilism come out as a form of consequentialism on Optimality?

No, it does not. We have already seen (in section 3) that beliefs do not come out justified on reliabilism simply on account of leading to a lot of epistemic value, including a *maximal* amount of such value. Moreover, a belief is not necessarily justified if produced by a process that maximizes the epistemic good. To see why, notice that the relevant notion of maximiza-

tion is a *relative* one: a process maximizes the epistemic good if it generates *more* epistemic value than any other process. That is compatible with the relevant process being unreliable, as will be the case in what we might call an *epistemically tragic world* where *all* processes are unreliable, and the maximally reliable one is simply slightly more reliable than the others—without being reliable in the *absolute* sense that the reliabilist is after. That means that, whether or not (1) comes out true on an Optimality reading, (2) does not.

## 6. Conclusion

To the best of our knowledge, sections 2 to 5 cover all of the criteria for consequentialism that are available in the literature. What we have shown is that there doesn't seem to be any plausible notion of consequentialism that has both (1) and (2) come out true, and that anyone giving the TOA against reliabilism thereby equivocates. Consequently, it is no surprise that (3) is false, nor should we expect further work on the part of the anti-reliabilist to reveal any problematic trade-offs for reliabilism. As such, our investigation suggests that the very strategy of identifying trade-off problems for reliabilism is misguided, if pursued in the manner that it has been pursued recently. While reliabilism is *a* kind of consequentialism, it is not of a kind that should lead us to expect trade-off problems.

*DePauw University*

*University of Kent, Canterbury*

## Notes

We are grateful to the participants at University of Hamburg's *Institutskolloquium* in philosophy, and to Arvid Bave in particular, for valuable comments on an earlier draft of this paper. We would also like to thank three anonymous reviewers for this journal for helpful comments. Research for this pa-

per was made possible by a generous grant from the Leverhulme Trust, for the project “Epistemic Consequentialism: Problems and Prospects”.

<sup>1</sup> Goldman (1979) is the seminal work in this area. See also Goldman (1986).

<sup>2</sup> Many reliabilists—including Goldman (1999)—embrace the stronger thesis that true belief not only has epistemic value but that it is also the sole bearer of intrinsic, epistemic value. Since we are here concerned with certain criticisms of reliabilism’s theory of rightness, and with its allegedly consequentialist nature in particular, as opposed to with its theory of value, we will not be offering a defense of the latter here. However, see Ahlstrom-Vij (2013) for a defense of the relevant type of epistemic value monism.

<sup>3</sup> Many thanks to an anonymous reviewer for raising this point.

<sup>4</sup> For a more exhaustive treatment of this case, see Ahlstrom-Vij and Dunn (2014).

<sup>5</sup> It is perhaps worth emphasizing this point. Denying that reliabilism faces trade-off *problems* is not to suggest that reliabilists don’t accept certain trade-offs. For example, Littlejohn (2012) objects to reliabilism with reference to how it classifies some beliefs in lottery propositions as justified on broadly consequentialist grounds, and moreover suggests that the lottery case is a trade-off case. For reasons discussed in Ahlstrom-Vij and Dunn (2014), we are sympathetic to the idea that lottery propositions can be justified, but also deny that, if the lottery case involves a trade-off, it involves a *problematic* trade-off. Moreover, the distinction between problematic and unproblematic trade-offs would have to be accepted by everyone. To maintain that all trade-offs are problematic would be to say that you can never trade off losses against gains, but that’s highly implausible. After all, someone who rejected all trade-offs would have to give up exercising, brushing their teeth, saving for retirement, and all other activities that involve short-term pain for long-term gain.

<sup>6</sup> We find further support for the idea that embracing Side-Constraints is what generates trade-off problems by considering the kind of epistemic decision theory (EDT) pioneered by Joyce (1998), and more recently developed by Leitgeb and Pettigrew (2010a, b). EDT is a *subjective* form of consequentialism, requiring the maximization of *expected* good. As such, it both satisfies Side-Constraints,

since it never sanctions realizing less rather than more expected good, and also condones problematic trade-offs (e.g., Greaves 2013).

<sup>7</sup> Thanks to an anonymous reviewer for raising this question.

<sup>8</sup> Further, though we know of no one who defends Immediate Consequence Maximization, there are extant views that are thought of as instances of epistemic consequentialism and yet do not take into account the causal consequences of belief states when evaluating those belief states. For a particularly clear statement of such a view, see Konek and Levinstein (forthcoming).

<sup>9</sup> Portmore's parenthetical remark about permissibility *in the knowledge-supposing sense* is not relevant to the general lesson we will be drawing here about maximization being sufficient but not necessary for permissibility.

## References

- Ahlstrom-Vij, Kristoffer. 2013. "In Defense of Veritistic Value Monism," *Pacific Philosophical Quarterly*, vol. 94, no. 1, pp. 19-40.
- Ahlstrom-Vij, Kristoffer, and Jeffrey Dunn. 2014. "A Defence of Epistemic Consequentialism," *Philosophical Quarterly*, vol. 64, no. 257, pp. 541-551.
- Bentham, Jeremy. 1789. *An Introduction to the Principles of Morals and Legislation* (Repr., Garden City: Doubleday, 1961).
- Berker, Selim. 2013a. "Epistemic Teleology and the Separateness of Propositions," *Philosophical Review*, vol. 122, no. 3, pp. 337-393.
- . 2013b. "The Rejection of Epistemic Consequentialism," *Philosophical Issues*, vol. 23, pp. 363-387.
- Brown, Campbell. 2011. "Consequentialize This," *Ethics*, vol. 121, no. 4, pp. 749-71.
- Chase, James. 2004. "Indicator Reliabilism," *Philosophy and Phenomenological Research*, vol. 69, no. 1, pp. 115-137.
- Firth, Roderick. 1981. "Epistemic Merit, Intrinsic and Instrumental," *Proceedings and Addresses of the American Philosophical Association*, vol. 55, no. 1, pp. 5-23.

- Foot, Philippa. 1985. "Utilitarianism and the Virtues," *Mind*, vol. 94, no. 374, pp. 196-209.
- Goldman, Alvin. 1979. "What Is Justified Belief?" in *Justification and Knowledge*, ed., George Pappas (Dordrecht, Holland: Springer), pp. 1-23.
- . 1986. *Epistemology and Cognition* (Cambridge, MA: Harvard University Press).
- . 2015. 'Reliabilism, Veritism, and Epistemic Consequentialism'. *Episteme*, vol. 12, no. 2, pp. 131-143.
- Greaves, Hilary. 2013. "Epistemic Decision Theory," *Mind*, vol. 122, no. 488, pp. 915-952.
- Howard-Snyder, Frances. 1994. "The Heart of Consequentialism," *Philosophical Studies*, vol. 76, no. 1, pp. 107-129.
- Jenkins, Carrie. 2007. "Entitlement and Rationality," *Synthese*, vol. 157, no. 1, pp. 25-45.
- Joyce, James. 1998. "A Nonpragmatic Vindication of Probabilism," *Philosophy of Science*, vol. 65, no. 4, pp. 575-603.
- Konek, Jason, and Ben Levinstein. Forthcoming. "The Foundations of Epistemic Decision Theory," *Mind*.
- Leitgeb, Hannes, and Richard Pettigrew. 2010a. "An Objective Justification of Bayesianism I: Measuring Inaccuracy," *Philosophy of Science*, vol. 77, no. 2, pp. 201-235.
- . 2010b. "An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy," *Philosophy of Science*, vol. 77, no. 2, pp. 236-272.
- Littlejohn, Clayton. 2012. *Justification and the Truth Connection* (Cambridge: Cambridge University Press).
- McCloskey, H. J. 1957. "A Non-Utilitarian Approach to Punishment," *Inquiry*, vol. 8, no. 1-4, pp. 239-255.
- McNaughton, David and Piers Rawling. 1991. "Agent-Relativity and the Doing-Happening Distinction," *Philosophical Studies*, vol. 63, no. 2, pp. 167-185.
- Mill, John Stuart. 1861. *Utilitarianism* (Repr. J. Bentham and J. S. Mill, *Utilitarianism and Other Essays*, London: Penguin, 1987).
- Nozick, Robert. 1974. *Anarchy, State, and Utopia* (New York: Basic Books).

- , 1981. *Philosophical Explanations* (Cambridge, MA: Belknap Press).
- Parfit, Derek. 2011. *On What Matters, Volume One* (Oxford: Oxford University Press).
- Percival, Philip. 2002. "Epistemic Consequentialism," *Proceedings of the Aristotelian Society, Supplementary Volumes*, vol. 76, pp. 121-151.
- Portmore, Douglas. 2007. "Consequentializing Moral Theories," *Pacific Philosophical Quarterly*, vol. 88, pp. 39-73.
- Rawls, John. 1963. "Constitutional Liberty and the Concept of Justice," *Nomos*, vol. 6, pp. 98-125.
- Sinnott-Armstrong, Walter. 2011. "Consequentialism," in *Stanford Encyclopedia of Philosophy*, ed., E. N. Zalta, available at: <http://plato.stanford.edu/archives/spr2014/entries/consequentialism/>.
- Slote, Michael. 1984. "Satisficing Consequentialism," *Proceedings of the Aristotelian Society*, vol. 58, pp. 139-163.
- Smith, Michael. 2009. "Two Kinds of Consequentialism," *Philosophical Issues*, vol. 19, pp. 257-272.
- Thomson, Judith Jarvis. 1976. "Killing, Letting Die, and the Trolley Problem," *Monist*, vol. 59, pp. 204-217.
- Williams, Bernard. 1973. "A Critique of Utilitarianism." in *Utilitarianism: For and Against*, eds., J.J.C. Smart and Bernard Williams (Cambridge: Cambridge University Press), pp. 77-150.