Reply to Goldman: Cutting Up the One to Save the Five in Epistemology

Selim Berker
Harvard University
sberker@fas.harvard.edu

[Penultimate draft of a paper that was eventually published in Episteme 12 (2015): 145–53; please cite that version.]

Abstract: I argue that Alvin Goldman has failed to save process reliabilism from my critique in earlier work of consequentialist or teleological epistemic theories. First, Goldman misconstrues the nature of my challenge: two of the cases he discusses I never claimed to be counterexamples to process reliabilism. Second, Goldman’s reply to the type of case I actually claimed to be a counterexample to process reliabilism is unsuccessful. He proposes a variety of responses, but all of them either feature an implausible restriction on process types, or fail to rule out cases with the sort of structure that generates the worry, or both.

I believe that a commitment to the epistemic analogue of consequentialist (or teleological) theories in ethics is pervasive in contemporary epistemology. These epistemic theories (i) take the good to be prior to the right in epistemology and (ii) explain epistemic rightness in terms of epistemic goodness via a promoting (or conducing) relation. Characterized thusly, epistemic consequentialism encompasses a wide array of currently popular theories, cutting across many of the familiar divides, including internalism vs. externalism, foundationalism vs. coherentism, and virtue- vs. belief-focused epistemology. But perhaps the most prominent instance of a consequentialist epistemic theory is process reliabilism, which I see as the epistemic analogue of satisfying hedonistic rule-utilitarianism (where instead of evaluating whether an action is based on a rule that does sufficiently well at promoting pleasure, we evaluate whether a belief is formed via a process that does sufficiently well at promoting true belief).

I also believe that this widespread commitment to consequentialism in epistemology is a mistake. My strategy for arguing against epistemic consequentialism has three stages (see Berker 2013a, 2013b). First, I consider certain counterexamples to extremely crude versions of epistemic consequentialism originally due to Roderick Firth (1981, 1998). Call these ‘Firth cases’. Second, I use the analogy with consequentialist theories in ethics to unearth the structural features that make Firth cases work the way they do. Third, I show how almost all forms of epistemic consequentialism sometimes allow a belief’s epistemic status to depend on those same structural features, thereby giving me a recipe for generating
counterexamples to even very sophisticated forms of epistemic consequentialism. Call the examples generated in this third stage ‘Firth-inspired cases’. They go beyond anything we find in Firth’s work, but, if I am right, they demonstrate the same underlying insight that led Firth to propose his own cases.

Since most epistemic consequentialists develop their theories in a way that allows them to deflect Firth cases, it is the second and third stages of my strategy that are crucial. The basic idea is to argue that all interesting forms of epistemic consequentialism condone (or mandate) the epistemic analogue of cutting up one innocent person in order to use her organs to save the lives of five people.¹ The difficult part is figuring out exactly what the epistemic analogue of cutting up the one to save the five consists in. This requires careful attention to the structure of Firth cases, so that we can use that structure to generate Firth-inspired counterexamples to the varieties of epistemic consequentialism one actually encounters in the literature.

In ‘Epistemic Teleology and the Separateness of Propositions’ (written in 2009), I postulated that the structural feature underlying Firth cases is this: the promotion of epistemic good with regard to one proposition is sacrificed in order to greatly benefit the promotion of epistemic good with regard to many other propositions. In slogan form: theories that run afoul of Firth’s original examples do so because they ‘ignore the separateness of propositions’. In ‘The Rejection of Epistemic Consequentialism’ (written two years later), I changed my mind and decided that although the structural feature of ignoring the separateness of propositions suffices to generate problems for many consequentialist theories in epistemology, there is an even more general structural feature subsuming that structural feature which allows us to generate problem cases even for those consequentialist epistemic theories which adhere to the separateness of propositions. I am still struggling to figure out the best way to characterize this more general structural feature. But that won’t matter for the purposes of this note: all of the theories I will be considering flout the separateness of propositions, so there will be no need to go beyond my initial diagnosis.

Alvin Goldman is the father of process reliabilism, so it is not surprising that he finds my criticism

¹ I refer here, of course, to a famous problem case for act-utilitarianism originally due to Judith Jarvis Thomson (1976: 206).
of epistemic consequentialism in general, and of process reliabilism in particular, unpersuasive. Goldman argues in ‘Reliabilism, Veritism, and Epistemic Teleology’ that the form of process reliabilism he favors can avoid my Firth-inspired attack. What exactly is process reliabilism? Here some terminology will help. Goldman (1979: 38) thinks of a process as ‘a functional operation or procedure, i.e., something that generates a mapping from certain states—“inputs”—into other states—“outputs”’. In the case of cognitive processes, the output states are all beliefs. On the input end, we can distinguish between belief-independent processes, which take no belief-states as inputs, and belief-dependent processes, which take at least some belief-states as inputs. We can also distinguish between process tokens and process types in the usual manner. It is process types, not process tokens, which are assessed for reliability, and there are two ways of doing this, depending on whether the process is belief-independent or -dependent:

A belief-independent process type is reliable iff its tokens tend to yield a sufficiently high ratio of true to false (token) output beliefs.

A belief-dependent process type is conditionally reliable iff its tokens tend, when their (token) input beliefs are all true, to yield a sufficiently high ratio of true to false (token) output beliefs.

Process reliabilism may then be defined recursively as follows:

process reliabilism:

i. If S’s (token) belief that \( p \) at \( t \) results (‘immediately’) from a belief-independent process token whose relevant process type is reliable, then S’s belief that \( p \) at \( t \) is justified.

ii. If S’s (token) belief that \( p \) at \( t \) results (‘immediately’) from a belief-dependent process token whose relevant process type is conditionally reliable, and if the other belief tokens on which this process token operated are themselves justified, then S’s belief that \( p \) at \( t \) is justified.

iii. Otherwise, S’s (token) belief that \( p \) at \( t \) is unjustified.

How do we determine the ‘relevant’ process type for a given process token? This, of course, is the dreaded generality problem for reliabilism.

Goldman’s reply to my anti-teleological argument consists in arguing that three of my cases are

---

2 Since I will not be discussing non-cognitive processes in this note, hereafter whenever I write ‘process’ I mean ‘cognitive process’.

3 What follows is almost exactly the formulation of process reliabilism that appears in Goldman 1979: 40–41, except that I have been explicit about the token vs. type distinction in the way proposed in Feldman 1985: 160–62.

4 Goldman himself noted this problem in Goldman 1979: 39, but since then it has been forcefully pressed by Earl Conee and Richard Feldman (see Feldman 1985; Conee and Feldman 1998; Feldman and Conee 2002; and Conee 2013). I agree with Conee and Feldman that the generality problem is an unsolved problem for process reliabilists, but I do not wish to rest my criticism of process reliabilism on that problem, since it does not generalize to all forms of epistemic consequentialism.
not counterexamples to process reliabilism: the grant-seeking scientist case (a Firth case whose specific formulation I take from Fumerton 2001), the sick mathematician case (another Firth case), and the prime numbers case (a Firth-inspired case customized to cause trouble for process reliabilism). Much of what Goldman says about these specific cases strikes me as correct. However, in restricting his reply to an assessment of these three cases, Goldman has failed to adequately address my anti-teleological strategy, for two reasons. First, I never claimed that Firth cases are counterexamples to process reliabilism. Second, Goldman’s argument that the prime numbers case (and others like it) do not present problems for process reliabilists is uncompeiling.

All Firth cases involve a seemingly unjustified belief being a causal means of promoting, sometime in the future, the acquisition of numerous true beliefs by the subject.5 Firth cases are not counterexamples to process reliabilism, since in the definitions of unconditional and conditional reliability, ‘yield’ should be interpreted to mean ‘immediately yield’. All that matters for an assessment of either type of reliability is the truth-ratio of the token belief-states directly brought about by token instances of the process type being evaluated; any other token belief-states causally downstream from those token belief-states are irrelevant. So I agree with Goldman that the grant-seeking scientist and sick mathematician cases are not counterexamples to process reliabilism. In fact, I explicitly said as much in my two articles.6

5 Or, at least, this is what Firth cases involve after we restrict the relevant epistemic end to true belief. (Firth himself discusses cases in which it is future knowledge, not merely future true belief, that is promoted.) In my work I call the view that the final ends in epistemology are true belief and the avoidance of false belief ‘veritism’, a term I say in Berker 2013a: 360 that I ‘adapt[ed]’ from Goldman. Goldman (2013: 139–40) takes issue with my using ‘veritism’ to cover process reliabilism and insists that, as he uses the term, it refers only to the complete consequentialist theory of social epistemology defended in Goldman 1999a. However, this does not fit with Goldman’s use of that term elsewhere in his work. As early as Goldman 1987: 122, he defines veritism as the view that ‘true belief [is] the ultimate epistemic aim’, a view which (a) is as much suited to individual epistemology as it is to social epistemology, and (b) is naturally seen as one component of process reliabilism (it specifies that theory’s account of final value, in my tripartite way of dividing up consequentialist theories). Similar comments apply to Goldman’s use of the term ‘veritism’ in Goldman 2001.

6 In Berker 2013a: 375, I wrote, ‘For this reason [namely, the fact that process reliabilists restrict the conducing relation to proximate causal means], none of the trade-off cases I have discussed so far pose a problem for the process reliabilist’. I say something similar in Berker 2013b: 374. Note that in both places I am referring not just to the original versions of the grant-seeking scientist and sick mathematician cases, but also to the tweaked versions of those cases in which we make the situations faced by the scientist and mathematician commonplace in their world. Thus it is simply a misreading on Goldman’s part to hold that I think the tweaked versions of those cases are counterexamples to process reliabilism due to the true beliefs causally downstream from the direct outputs of the processes in question.

(This last qualification is crucial when it comes to the case of the sick mathematician. All I am claiming here is that the sick mathematician is a not a counterexample to process reliabilism in virtue of the true beliefs causally brought about by the mathematician’s belief that she will recover. Depending on how we individuate process types, it may well be the case that the sick mathematician is a counterexample to process reliabilism in virtue of the fact that the mathematician’s belief that she will recover has a causal impact on its own truth. However, I won’t be pressing that point in the current note. For discussion, see Berker 2013b: 376–77.)
But process reliabilists are not so easily off the hook. Note the structure of Firth cases: they involve forgoing the promotion of true belief with regard to one proposition in order to greatly benefit the promotion of true belief with regard to many other propositions. In Firth’s original cases, the item doing the forgoing and promoting is a belief token. However, that same underlying structure can also apply to cases in which something else, such as a process type, is sacrificing its performance with regard to one proposition in order to promote its performance with regard to many other propositions. For this reason, process reliabilism sometimes allows the structure at work in Firth cases to be normatively relevant, and when it does, I claim, it reaches an incorrect verdict. My prime numbers example is supposed to be one (but only one) case of this sort.

Here was my initial description of the case: ‘Suppose that whenever I consider whether a given natural number \( n \) is prime, I form the following belief: \(<n \text{ is not prime}>\).’ The general structure of the case—which I will call ‘structure \( S \)—was intended to be as follows:

\[
\text{structure } S:
\]

a. There is a belief-independent process, \( X \), which outputs beliefs in a range of propositions, \( R \).

b. There is a partition of \( R \) into two subsets, \( R_1 \) and \( R_2 \), such that:
   - when \( X \) outputs a belief in a proposition in \( R_1 \), that belief is always false;
   - when \( X \) outputs a belief in a proposition in \( R_2 \), that belief is always true;
   - \( R_1 \) is much smaller than \( R_2 \).

c. As a result, \( X \) is extremely reliable.

Why are cases with structure \( S \) problematic for process reliabilism? Let us refer to the propositions in \( R_1 \) as ‘the few’ and the propositions in \( R_2 \) as ‘the many’. Intuitively,\(^8\) when a subject uses \( X \) to form a belief in one of the few, that belief is not thereby justified: \( X \)’s poor performance with regard to the few is not redeemed by its excellent performance with regard to the many. But process reliabilism appears to yield

\(^7\) Berker 2013a: 375. The case reappears in Berker 2013b: 374–75.

\(^8\) Here I suppress the type/token distinction, for readability’s sake.

\(^9\) I am using ‘intuitively’ here as a placeholder for the proper means, whatever it is, by which we come to correct epistemic verdicts about particular-case scenarios. Skeptics of intuition can feel free to replace that word with a reference to their favorite alternative. (What if there is no way of properly coming to correct epistemic verdicts about particular-case scenarios? Well, then epistemology is doomed, and this entire exchange between me and Goldman is moot.)
the verdict that such a belief is justified.

Goldman is not moved by this case. He presents a scattershot of possible responses. Two of these we can immediately set aside. First, Goldman (in effect) points out—quite correctly—that (c) only follows from (a) and (b) in S if we make an additional assumption, such as:

d. X tends to output beliefs in a given proposition in \( R_1 \) about as often as it tends to output beliefs in a given proposition in \( R_2 \).

Moreover, he doubts that (d) is true of the case as I describe it. Goldman (2015: 141) writes, ‘In all likelihood, people who use this process type apply it exclusively or primarily to relatively small integers. Within that class of numbers, however, the truth-ratio of the process will not be very high, because many small integers are prime’. But this is at most an objection to my particular attempt to describe a case with features (a), (b), and (c), not to the very possibility of there existing a case with that structure. So let us stipulate that instead of the process being one that yields belief in propositions of the form \(<n \text{ is not prime}>\) for arbitrary values of \( n \), it only yields belief in propositions of that form for values of \( n \) that are at least three digits.¹⁰ It is easy enough to imagine a case of that sort for which assumption (d) is true. Second, Goldman notes—quite correctly—that in many of his writings on process reliabilism he makes a distinction between belief-forming processes and belief-forming methods which my example appears to neglect. In Goldman’s words:

> ‘Processes’ are basic psychological processes, roughly, wired-in features of our native cognitive architecture. ‘Methods’ are learnable algorithms, heuristics, or procedures for forming beliefs, such as procedures that appeal to instrument readings, or statistical analyses.¹¹

Process reliabilism, Goldman insists, is only intended to be an account of the justifiedness of beliefs formed entirely on the basis of processes; for beliefs formed at least partially on the basis of methods, the account of justifiedness is allegedly more complicated, and in particular involves assessing whether the subject acquired the relevant methods in an appropriate manner.¹² Moreover, the prime numbers case, as I described it, may seem to involve a consciously applied rule for forming beliefs about the primeness of

---

¹⁰ And, we might add, at most thirty digits, to obviate any worries concerning infinities.


¹² Thus on this proposal, we should interpret ‘results’ to mean ‘entirely results’ in clauses (i) and (ii) of my statement of process reliabilism, and clause (iii) needs to be reformulated.
natural numbers, and hence may seem to feature a belief-forming method, not a belief-forming process. But, again, this is at most an objection to my particular attempt to describe a case with structure S, not to the very possibility of there existing a case with that structure. We can easily change the case so that it features a subject with a ‘wired-in’ feature of her ‘native cognitive architecture’ that allows her to form beliefs about the primeness of natural numbers in the way specified. And when we do, I claim, that does not affect our judgments about the case: it is still implausible to hold that beliefs in the few formed on the basis of this process are justified.\textsuperscript{13}

Goldman has a third reply to the prime numbers example which is more promising than the other two.\textsuperscript{14} This reply involves, in effect, denying that the process type specified in the example is an \textit{epistemically relevant} one. Goldman notes that in his first article defending process reliabilism, he suggested that relevant process types should be restricted to those that are ‘content-neutral’ (Goldman 1979: 39). This restriction was designed to rule out processes such as the following: \textit{inferring \textless p\textgreater \text{from} \textless The Pope asserted that p\textgreater}.\textsuperscript{15} And, Goldman insists, that same restriction rules out the process in my prime numbers case.

I am not convinced. I will now argue that depending on how we interpret the proposed restriction to content-neutral process types, either that restriction is extremely implausible (call this ‘the plausibility problem’), or it fails to rule out problem cases with the relevant structure (call this ‘the effectiveness problem’), or both.\textsuperscript{16}

\begin{footnotesize}
\begin{enumerate}
\item Objection: The new version of the prime numbers case features a process type that is not instantiated in the actual world. But ‘the domain we select for determining a process-type’s reliability is the actual world (plus, perhaps, a class of relevantly nearby possible worlds)’ (Goldman 2015: 141). So process reliabilists are not forced to say that the process type in this new version of the case is reliable.

\textit{Reply:} We should distinguish between the following two claims: (1) we need to look to the actual world (and certain nearby worlds) when selecting a relevant process type, and (2) when determining whether an already selected process type is reliable, we do so by seeing how reliable it would be if embedded in the actual world (and certain nearby worlds). Presumably in this quotation Goldman is only putting forward the second claim. Otherwise process reliabilism has no way of assessing beliefs formed on the basis of process types that are not actually instantiated. And if process reliabilism has no way of assessing beliefs formed on the basis of process types that are not actually instantiated, then so much the worse for process reliabilism.

\item Although Goldman presents this third reply in the same paragraph as his reply involving an appeal to the process/method distinction, I take it that these two replies are logically distinct from one another.

\item That is, a process which takes as input a belief in the proposition \textless The Pope asserted that \textless p\textgreater and yields as output a belief in the proposition \textless p\textgreater. Such a process is problematic if we agree that a belief formed (or sustained) on the basis of that process is not justified, even if, as a matter of fact, the Pope happens to be infallible.

\item A similar objection applies to David Copp’s (2014: 88–90) attempt to defend his own brand of epistemic teleology against my criticisms by appealing to an optimality condition together with what he calls ‘the learnability constraint’ and ‘the generic
\end{enumerate}
\end{footnotesize}
It is very unclear what content-neutrality comes to. In his initial discussion of the content-neutral restriction, Goldman (1979: 39) defined content-neutrality as follows:

*proposal #1:* A process is content-neutral iff it ‘admit[s] as input beliefs (or other states) with any content’.

When content-neutrality is understood in this way, the resulting restriction on process types faces a plausibility problem. In particular, many of Goldman’s own examples of relevant process types would be ruled out by this version of the content-neutral restriction, including ‘sight’, ‘hearing’, ‘ignoring contrary evidence’, ‘beliefs based on inference from observed samples’, and ‘accepting the reports of others’.17 None of these examples involve input states with arbitrary contents. This first way of interpreting the content-neutral restriction also faces an effectiveness problem. Because it proposes no constraints on the content of output beliefs, the restriction fails to rule out the process type in the prime numbers case. (My description of that case is silent on the content of the inputs to the process described, if indeed those inputs even have content.) To avoid this last problem, we might switch to

*proposal #2:* A process is content-neutral iff it admits as input beliefs (or other states) with any content and admits as output beliefs with any content.

This alteration helps with the effectiveness problem, but now we have a *drastic* plausibility problem. Pretty much the only process type that satisfies the content-neutral restriction under this interpretation is deductive inference.18

Goldman (2015: 141) has now changed his definition of content-neutrality to

*proposal #3:* A process is content-neutral iff it ‘take[s] [states with] many different contents as inputs (and as outputs)’.

Presumably ‘many different’ refers not just to the *number* of different contents—the process in my prime numbers case does very well by that count—but also to the *diversity* of those contents. Under this interpretation the content-neutrality restriction is less implausible (though perhaps misnamed: ‘content-flexible’ strikes me as a more accurate term than ‘content-neutral’ for the notion being glossed here).

---

17 The first three of these I take from Goldman 1992: 158, the fourth from Goldman 1979: 38, and the fifth from Goldman 1999a: 129.

18 And maybe not even that, if it turns out that deductive reasoning in humans is domain-specific, a possibility about which Goldman wishes to remain uncommitted (see Goldman 1999b: 17–19).
However, this version of the restriction is beset by an effectiveness problem. It is easy enough to tweak the prime numbers case so as to make process X content-neutral in the sense proposed: just keep adding extra propositions which are not of the form \(<n \text{ is not prime}>\) into \(R_2\) (the set of propositions that X always gets correct) until a desired level of diversity is achieved. This change does not, I claim, alter our verdicts about the case.

Thus I am not convinced by the three replies we have considered to my prime numbers example. More generally, we can separate two issues with regard to that case. First, we may ask, ‘Does process reliabilism allow cases with structure S to exist?’ Second, we may ask, ‘If cases with structure S exist, would they be a problem for process reliabilism?’ Goldman seems to be conceding that the answer to the second of these questions is ‘Yes’. His strategy is to resist the example by attempting to answer ‘No’ to the first question. But this strikes me as a not-very-promising strategy. When \(R\) ranges over necessary propositions, cases with structure S are a dime a dozen—as we have seen, from the failure of Goldman’s various replies to rule out cases with that structure. Moreover, if we let \(R\) range over contingent propositions and change ‘always’ to ‘almost always’ in clause (b), then cases with the resulting structure are also a dime a dozen. Finally, we can change ‘belief-independent’ to ‘belief-dependent’ in clause (a) and make accompanying changes in clause (b) to generate yet more cases with a problematic structure. So really structure S is a template for a vast number of ways in which a process type might forgo success with regard to certain propositional outputs in order to achieve great success with regard to many other propositional outputs. We can imagine cases of this sort involving not just mathematical processes, but also perceptual processes, introspective processes, inferential processes—the list goes on. I think it is striking how resistant we are to saying that such cases involve a ‘sacrifice for the greater epistemic good’.

\[19\] What about Goldman’s fourth proposed response, in which he appeals to what he now calls ‘the approved-list version of reliabilism’—and what he used to call ‘two-stage reliabilism’ (Goldman 1992, 1999b)—in order to reach ‘the correct verdict’ about the prime numbers case, namely that a belief in \(<7 \text{ is not prime}>\) on the basis of process X is not justified (Goldman 2015: 142)? In offering this response, Goldman is, in my opinion, misapplying his own theory. What Goldman and I are debating here is the viability of process reliabilism as a theory of what in fact makes a belief justified (i.e. as a theory in ‘normative scientific epistemology’, as Goldman [1992: 156] calls it). But in his writings, Goldman only offers approved-list reliabilism as a theory of how the folk make judgments about whether a belief is justified (i.e. as a theory in ‘epistemic folkways’ or ‘descriptive scientific epistemology’, as Goldman [1992: 135–56] calls it)—which is all for the better, since approved-list reliabilism is not remotely plausible as a theory of what in fact makes a belief justified. So appealing to approved-list reliabilism is not a way of establishing the ‘correct verdict’ which Goldman hopes to secure; at most it is a way of establishing that such a verdict is part of our folk theory of justified belief. Moreover, I make no claims in my work about whether epistemic consequentialism accurately captures our folk theory of justified belief, since in my opinion folk epistemology is of limited philosophical interest.
in the way in which an act-consequentialist might deem a cutting up of the one to save the five as a sacrifice for the greater moral good. This judgment of ours about the justificatory status of beliefs formed (or sustained) on the basis of processes with structures similar to S is not, I claim, merely a folk verdict, but rather is a persistent considered judgment that survives reflective scrutiny. And it is a judgment which Goldman seems to share. I conclude that Goldman has not successfully shown that process reliability avoids my charge that the view illicitly ignores the separateness of propositions.  

REFERENCES


20 I thank all of the other participants of the Tenth Anniversary Episteme Conference in San Jose, Costa Rica, for their questions and suggestions, and I thank Tyler Doggett, Emily McWilliams, Michael Rabenberg, Gustavo Ribeiro, and Susanna Siegel for comments on a previous draft of this article. But most of all, I thank Alvin Goldman for the inspiration and impetus which his work has provided me over the years, despite our differences in epistemological outlook. There is a distinctive mix of clarity, rigor, creativity, and forcefulness of argument in Goldman’s philosophical writing that I hope one day to achieve in my own. Work on this article was generously supported by fellowships from the American Council of Learned Societies and the Radcliffe Institute for Advanced Study.


