

Belief Ownership without Authorship:
Agent Reliabilism's Unlucky Gambit against Reflective Luck

Benjamin Bayer
September 1st, 2014

Abstract: This paper examines a persuasive attempt to defend reliabilist theories of justification against the influential “reflective luck” objections popularized by Laurence Bonjour and Keith Lehrer. In a series of papers, John Greco and Daniel Breyer have argued that an agent reliabilist version of externalism has the resources to explain why subjects who are the lucky beneficiaries of reliable processes such as clairvoyance and implanted devices do not count as justified after all. They claim a necessary condition of the “subjective” aspect of agent-reliable justification is the subject’s ownership of the belief, and that subjects in these reflective luck cases do not own their beliefs in the relevant way. This paper considers an additional example of reflective luck described by Jennifer Duke-Yonge that seems to run afoul of the agent reliabilist strategy, and shows how it can be further developed in response to Daniel Breyer’s recent criticisms of Duke-Yonge.

Introduction: reflective luck objections to reliabilism

Reflective luck objections to reliabilism describe cases in which subjects possess true beliefs that appear to be justified according to reliabilist externalist criteria of justification, but which are only accidentally true from the perspective of the subject’s reflective awareness. These beliefs are accidentally true or true as a matter of luck in the minimal sense of being external to a subject’s conscious choices, thoughts and actions.¹ Many think that such beliefs are obviously unjustified, and that these cases serve as compelling counterexamples to reliabilism.

A classic reflective luck counterexample is Laurence Bonjour’s example of Norman the clairvoyant. In Bonjour’s account Norman a) is a completely reliable clairvoyant on specific topics, b) has no evidence for or against the possibility of clairvoyance in general, or for or against his possession of the power in particular, but c) believes that the president is in New York City as a result of possessing his power (Bonjour 1985 :41). Bonjour thinks that it would be “epistemically irrational and irresponsible, and thereby unjustified” for Norman to accept such a belief. If the example of Norman seems too fantastic, a second prominent example of reflective luck is not so easily dismissed. Keith Lehrer’s example of Mr. Truetemp does not so obviously appeal to scientifically disreputable faculties. Truetemp a) has a surgically implanted “tempucomp” that generates reliable temperature thoughts, b) has no evidence for or against his possession of the tempucomp, and c) believes some of these thoughts (though he is puzzled about why he thinks so obsessively about temperature) (Lehrer 2000: 187). Lehrer like Bonjour thinks it would be unjustified for Truetemp to form temperature beliefs due to the output of this device.

¹ I borrow this understanding of reflective luck from John Greco (2006: 15) via Breyer (2010: 134).

In both such cases, an agent's beliefs seem to be wholly unjustified even though they meet reliabilist criteria for justification, because their accuracy and reliability seems to be a matter of good luck for the agent in question. In this paper I will argue that recent attempts to defend reliabilism against these counterexamples by refining the reliabilist account along agent reliabilist lines do not succeed.

The agent reliabilist criteria of justification vs. reflective luck

Recently, John Greco (1999, 2004) has argued for a form of virtue-based or agent reliabilism that seeks to accommodate the reflective luck counterexamples. His position stresses that a belief's justification has both an "objective" and "subjective" component, such that it is not enough that the agent's beliefs be the result of a reliable process (the "objective" component): their reliability must itself be the result of an agent's intellectual virtue (the "subjective" component). Knowledge, in this regard, is a form of "success through virtue" (Greco and Breyer, 2008, 174). This formulation is thought to imply that the relevant cases of reflective epistemic luck can be shown to lack justification according to reliabilism after all, because agent reliabilism treats only a subset of beliefs produced by reliable processes as justified.

Greco and Breyer have developed a series of criteria aimed at unpacking the idea of success through virtue. To my knowledge, they have not presented all of the criteria in a single formulation. Rather than recount their presentation as it has unfolded in a series of exchanges, I will endeavor here to present it all at once, to show as quickly as possible how it is supposed to address the reflective luck problem, and then to show why I think it does not succeed. Here, then, is the set of successive criteria that together unpack the full idea of "success through ability":

- (1) A subject S's belief P is justified *iff* S believes the truth with respect to P reliably (is *objectively* justified) because S's belief that P is produced by intellectual virtue (is *subjectively* justified). (Breyer and Greco 2008, 173-174)
- (2) S's belief that P is produced by intellectual virtue (is subjectively justified) *iff* the belief that P is produced by dispositions that S manifests when S is motivated to believe the truth. (Breyer and Greco 2008, 174)
- (3) The belief that P is produced by dispositions that S manifests when S is motivated to believe the truth *only if* the belief that P results from dispositions that are well-integrated with other dispositions that S manifests when motivated to believe the truth. (Breyer and Greco 2008, 175)
- (4) The belief that P results from dispositions that are well-integrated with other dispositions that S manifests when motivated to believe the truth, *only if*
 - (a) the belief that P is not subject to any defeaters to which S has access (Breyer and Greco 2008, 175)
 - (b) S owns the belief that P. (Breyer and Greco 2008, 178)

- (5) S owns the belief that P *only if* the belief that P results from a cognitive mechanism for which S has taken responsibility, which S does *only if*:
- (a) S recognizes herself as the source of her beliefs, such as perceptual beliefs,
 - (b) S accepts that she is fairly credited with having certain beliefs, and
 - (c) S bases her beliefs about herself on appropriate grounds (Breyer and Greco 2008, 181)²

Notably, the view of belief ownership expressed in (5) is *historical*: past acts of taking responsibility in the manner above allow for a presently owned belief. Recalling this will be important for our pending evaluation of this position.

Now (3) through (5) only state necessary conditions for a belief's being produced by dispositions that S manifests when S is motivated to believe the truth, but that is all Greco and Breyer need here: if a given belief fails to be owned by S, it cannot be justified. This is how they intend to deal with the relevant cases of reflective luck. I say the "relevant" cases, because some cases can be dealt with without appealing to (4)(b) or (5). Some of these cases might also be ruled out by (4)(a), the "no defeaters" condition. Indeed Greco and Breyer propose that we might interpret cases like Norman the clairvoyant in one of three ways. In the case of Norman I (unlike the case presented by BonJour), Norman actually does have evidence against the possibility of clairvoyance and is concerned to respect counterevidence (he shares our defeater system): if he still believes the president is in New York City, his belief is unjustified according to 4(a) above. But BonJour's presentation is ambiguous between two additional interpretations of the Norman who does not share our defeater system: Norman II shares the system and still owns the belief because he is a very different creature from us (his clairvoyance is like visual perception is to us); Norman III shares the system but does not own the belief because it stems from a subpersonal system for which he does not take responsibility. Hence a Norman II may still be seen as having a justified clairvoyant belief, but not Norman III.

For the time being I will leave aside the question of whether we should so easily assume that Norman II really does have a justified belief, and focus on Norman III. In his (2010) elaboration on the taking responsibility account, Breyer elaborates on why a clairvoyant like Norman III *does not* own his belief. Norman III is more like a "belief machine" than a real believer: his clairvoyant belief is not really *his* except in the weakest of senses: it's not anyone else's, but it also doesn't flow from the exercise of his own agency. Because it results from a merely subpersonal mechanism, it comes to him "out of the blue" and is "alien" to him. Because it is not his belief by (5), it is not well-integrated into the other dispositions he manifests when he is motivated to believe the truth by (4), and so by (3) and (2), his belief is not justified.

² Breyer and Greco propose this criteria as parallel to Fischer and Ravizza's (1998) account of moral responsibility which is cashed out partially in terms of taking responsibility for reasons-responsive mechanism.

Two counterexamples to agent reliabilism

Even still, it is not entirely clear that agent reliabilists can use their criterion of belief ownership to show how cases of reflective luck can be counted as non-justified. Jennifer Duke-Yonge (2013) proposes that there is a salient case of reflective luck that cannot be counted as an alien belief by Greco and Breyer's account and so cannot be counted as non-justified by this standard. This time, however, Duke-Yonge points to a new version of the Truetemp counterexample.

Duke-Yonge proposes a "sequel" to the original Truetemp story (let's call him Truetemp II). She asks us to suppose that Truetemp II discovers that he has the tempucomp in his brain, and comes to know that it gives him reliably true beliefs about temperature. But Truetemp II decides to have the tempucomp removed, and his doctors tell him that after the device is removed, he will experience random temperature thoughts for some time. Truetemp II then experiences temperature thoughts after his procedure, and dismisses them as the random after-effects he had been expecting. But in another twist, it turns out that the doctors lied and did not remove the tempucomp from Truetemp II's head, so these temperature thoughts are still reliably true. Duke-Yonge's counterexample focuses on the fact that belief ownership, according to Greco and Breyer's agent reliabilism, is supposed to be a *historical* fact, generated by past acts of taking responsibility. Even though Truetemp II may explicitly disavow ownership of his temperature beliefs at this point, the fact remains that his beliefs are issuing from a belief-forming mechanism for which he *has* taken ownership in the past. Since Breyer does not require any reflection on the source of these beliefs as a condition for owning them, it seems that his version of agent reliabilism does not allow him to classify as alien *any* of Truetemp's temperature beliefs. And yet it seems they clearly are.

Duke-Yonge suggests that agent reliabilists should appeal instead to a conception of belief *authorship*, which she illustrates using real world examples of "thought insertion" problems exhibited by schizophrenics. Schizophrenics will sometimes believe that thoughts that are introspectively accessible to them are not their own. They do not "own" these beliefs, she suggests, because they neither self-ascribe these beliefs nor are their authors, i.e. they do not endorse them on the basis of reasons. But by the same token, Truetemp II should not count as owning his beliefs either. He neither self-ascribes the temperature beliefs nor is their author in the sense that he does not endorse them on the basis of reasons. However, both Truetemp I and Norman III may self-ascribe for all we know. If we still think their temperature or clairvoyant beliefs are not their own, then authorship has to be the decisive necessary condition for belief ownership that is missing in these cases.

Breyer (2013) responds by noting that on his view, belief ownership is a necessary condition ((4)(b)) but not a sufficient condition for subjective justification. Another necessary condition is not being subject to defeaters ((4)(a)), and if Truetemp II remembers his doctors' warning that he would experience random temperature thoughts, he is aware of a defeater for his belief. So the agent reliabilist does have a way of showing that Truetemp II's temperature beliefs are not justified. Furthermore, Breyer gives at least one reason for thinking that Truetemp II does not own his belief after all, even on his "taking responsibility" account. He suggests that because Truetemp II explicitly denies ownership of the belief, he no longer sees himself as the source of the belief and so is no longer taking responsibility for the belief. So the necessary condition in (4)(b) may be relevant after all.

I will now argue we can very easily revise the Duke-Yonge counterexample to show that there is still a problem for Breyer and Greco's account. Since we are already in the realm of a science fiction example, we could imagine that when Truetemp (let's call him Truetemp III) undergoes his procedure, the scientists, instead of removing the device, remove a number of his memories. They remove his memory of having discovered the reliability of his temperature beliefs along with the memory of his decision to have the procedure to remove the device and of their warnings concerning random beliefs. This means that Truetemp III no longer is aware of defeaters to his beliefs. His beliefs cannot be counted as non-justified according to (4)(a). If we suppose that he accepts them in a lazy way anyway, can we say that his belief is not justified simply on the grounds that he does not own the belief, according to (4)(b)? The problem is that the historical account of belief ownership understood in terms of taking responsibility implies that this Truetemp *does* own his temperature beliefs: it remains true that at one time in the past Truetemp took responsibility for these temperature beliefs (when he discovered that they were reliable, a fact he has now forgotten). So it seems we have a case fulfilling all of the agent reliabilist's conditions for subjective justification, and yet Truetemp III's beliefs seem to be neither genuinely his own nor justified.

Objections

Breyer could respond that (4)(a) and (4)(b) are only necessary conditions of having well-integrated or justified beliefs. That Truetemp III fulfills both does not imply that he has well-integrated or subjectively justified beliefs, because it's not been suggested that these two necessary conditions are jointly sufficient. But Breyer and Greco have not yet given any additional necessary conditions that one has to fulfill for one's beliefs to count as subjectively justified, and Truetemp III's beliefs *do* fulfill the objective justification requirement (they are reliable). It seems reasonable to regard the fulfillment of all of the necessary conditions as jointly sufficient in the absence of any additional necessary conditions. Breyer and Greco owe us an additional necessary condition if this counterexample is to be dismissed.

One could object that there are some obvious necessary conditions one could add to Breyer and Greco's account to explain why Truetemp III does not own his belief after all. For instance, one could adopt Duke-Yonge's authorship account, which she herself offers as a favorable to agent reliabilism. Interestingly, however, Breyer argues that the authorship requirement is more demanding than the historical account. He notes, for example, that it could not account for beliefs justified by perceptual knowledge—and presumably it might not count Norman II as knowing in the way that reliabilists think he should. In other words, it seems that the authorship account is far too internalistic for reliabilists.

One could also adopt a current self-ascription condition without adopting an authorship condition, and so claim that even if Truetemp III doesn't disavow the belief as his, he also hasn't positively claimed it as his either, and so he does not own his belief. One problem here is that there is at least a minimal sense in which he must self-ascribe the belief: it is part of the thought experiment (as it was with Norman III) that Truetemp III still *accepts* the belief in a lazy way. If acceptance involves at least a minimal form of self-ascription, then Truetemp III does still own the belief in every way the agent reliabilists have proposed.

But if a more robust kind of self-ascription is claimed to be necessary, we can then raise the serious question of whether or not this is consistent with externalism. Externalists usually emphasize the historical rather than the synchronic nature of the properties necessary for justification because they see continuous awareness as too demanding to account for all of the beliefs we take to be justified. For example Goldman (1999, 2009) accuses internalism of being incapable of solving the "problem of forgotten evidence," wherein (e.g.) a subject reads that broccoli is healthy in a reliable newspaper, and later forgets where she read it, but retains the belief. Goldman contends that only a preservative, historical account of memorial justification, rather than a synchronic account, can account for the justification of such a belief. It is very easy to think of Truetemp III's case as analogous to cases of forgotten evidence. If robust and synchronic self-ascription were necessary for the continuous ownership of beliefs, hardly anyone would own their beliefs, since we are not constantly self-ascribing all of our beliefs.

A final objection is that an internalist theory would fare no better in this regard, and so these objections should not be seen as decisive against the agent reliabilist's defense of externalism. Although I cannot develop it further here (because I have argued for it elsewhere), I do not think this is true. I think we can develop an authorship account of belief-ownership that is (by Breyer's own admission) quite consistent with internalism but which rules out Truetemp III's (and Norman III's) beliefs as neither his own nor justifiable. And I think there are good reasons to think that this account would not be too demanding, provided that what it would take to see oneself as the author of a belief in virtue of having an awareness of the good reasons for this belief would not be too demanding either. I have argued elsewhere

that perceptual justification involves a form of epistemic perspective to which ordinary subjects have access in large part because it is pre-conceptual and so not subject to epistemic regress.

Although I cannot develop the point here, the present line of thought does suggest another concern I have about the agent reliabilists' evaluation of Norman II: even if Norman II in some sense sees himself as the source of his reliable clairvoyant beliefs, I do not think this is adequate to count them as justified. If one does not know of their reliability, one could only see oneself as the source of these beliefs by inventing a rationalization for their acceptability. In such a case, it is not obvious how their justifiability would even be consistent with the agent reliabilist formula of being successful through virtue, for engaging in rationalization seems to be no intellectual virtue. But this is a response best developed elsewhere.

Bibliography

- Bonjour, L. (1985). *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.
- Breyer, D. (2010). Reflective Luck and Belief Ownership. *Acta Analytica*, 25, 133-54.
- Breyer, D. (2013). Ownership, Agency, and Defeat. *Acta Analytica*, 28, 253-256.
- Breyer, D., & Greco, J. (2008, January). Cognitive Integration and the Ownership of Belief: Response to Bernecker. *Philosophy and Phenomenological Research*, 76(1), 173-184.
- Duke-Yonge, J. (2013). Ownership, Authorship and External Justification. *Acta Analytica*, 28, 237-52.
- Goldman, A. (1999). Internalism Exposed. *Journal of Philosophy*, 96(6), 271-293.
- Goldman, A. (2009). Internalism, Externalism, and the Architecture of Justification. *Journal of Philosophy*, 106(6), 1-30.
- Greco, J. (1999). Agent Reliabilism. *Philosophical Perspectives*, 13, 273-96.
- Greco, J. (2004). Virtue Epistemology. *The Stanford Encyclopedia of Philosophy* (October 2004). (E. Zalta, Ed.) Retrieved from <http://plato.stanford.edu/entries/epistemology-virtue/>
- Greco, J. (2006). Virtue, Luck and the Pyrrhonian Problematic. *Philosophical Studies*, 130, 9-34.
- Lehrer, K. (2000). *Theory of Knowledge* (2nd ed.). Boulder: Westview Press.