

Penultimate version; final version in *Philosophical Studies* 134 (2007). Please feel free to email me for a pdf of the published version: david.christensen@brown.edu

### Three Questions about Leplin's Reliabilism

David Christensen

University of Vermont

Jarrett Leplin's paper is multifaceted; it's very rich with ideas, and I won't even try to touch on all of them. Instead, I'd like to raise three questions about the paper: one about its definition of reliable method, one about its solution to the generality problem, and one about its answer to clairvoyance-type objections.

1

The first question I would like to raise concerns the definition of a reliable method for producing or sustaining true beliefs. Leplin's definition specifies that a reliable method "wouldn't produce or sustain false beliefs in normal conditions." It is important in understanding this definition that "would not produce or sustain false beliefs" doesn't mean "would not often," or "would not as a general rule." It means "would not ever." The idea is that reliability is not to be understood in the statistical sense of yielding a high ratio of true beliefs to false ones. Rather, a reliable method is one that would never produce or sustain false beliefs--as long as conditions were normal.

My initial worry here is that this sets the standards for reliability too high. But whether that is so will depend, of

course, on what "normal conditions" are taken to be. They are intended to be conditions typical of situations in which the method is applicable, and conditions that agents naturally presuppose to obtain in using the method. Let us look at how this notion will be applied in a couple of examples of intuitively reliable methods.

Consider first "trusting the Encyclopedia Britannica". What are normal conditions for this method? It seems to me that the conditions that typically obtain, and which one naturally presupposes to obtain, when one uses the Encyclopedia Britannica clearly include the following: the Encyclopedia's articles are written by acknowledged experts in the relevant fields, and these articles are subject to careful and rigorous fact-checking. Yet it seems clear that these conditions allow the possibility of occasional error. Normal conditions must include more, if trusting the Encyclopedia is to count as reliable.

Another intuitively reliable method of forming beliefs is facial recognition of people one knows. Consider identifying an acquaintance 30 feet away on the basis of recognizing her face. It seems to me that the conditions that typically obtain, and which one presupposes to obtain, in such cases include: adequate lighting; the absence of disguises, hallucinations, or identical twins; and one's own good memory for faces. But these conditions

clearly are not sufficient to render the method infallible when they obtain.

But what would one add to normal conditions in these cases, to secure the reliability of the relevant methods? It might well be claimed that one condition that typically obtains, and that one presupposes to obtain, when one believes that P on the Encyclopedia's say-so, is simply that the Encyclopedia is correct about whether P. Although this may not be a very natural reading of "conditions," it would clearly secure the reliability of the method: on this understanding of normal conditions, trusting the Encyclopedia would, trivially, be an infallible method--under normal conditions. Analogously, it might be suggested that normal conditions for recognizing acquaintances by sight would include correct identification.

But allowing the normal conditions for using a method to include the method's delivering a true belief would bring problems of its own. If we then took normal conditions to be just those which agents presuppose to obtain, then trusting a Ouija board would come out as infallible under normal conditions, given that the user presupposed that the Ouija board got things right. Now that sort of example would be eliminated, if we also require that normal conditions typically or characteristically do obtain when the method is applicable. But the notion then threatens to collapse into a purely statistical notion of

reliability. Given that a method typically or characteristically does yield true beliefs, it will count as infallible under normal conditions. Thus it seems to me that defining reliability as infallibility under normal conditions threatens to yield a notion of reliable method that is either too demanding, or which reduces to a statistical notion of reliability.

Now I should say that it's not obvious that a statistical understanding of reliability would be detrimental to Leplin's wider purposes. For example, such a notion is compatible with his desideratum that a reliable method not guarantee the production or sustaining of false beliefs. Suppose I form beliefs about tickets in big lotteries that they will lose. Even if this is counted as reliable method, it remains possible that it never lead me astray even if I use this method many times a day for the rest of my life.

However, embracing a statistical understanding of reliability may sit less well with a different principle Leplin endorses: that justification is transmissible by truth-preserving inferences. The method of believing large conjunctions of statistically reliably produced beliefs is not itself a statistically reliable method. My own view is that this is no loss--that the preface and related cases show that the correct account of justification will not allow unrestricted transmission of justification over truth-preserving multi-premise inferences.

But I realize that my sanguine attitude toward rejecting the conjunction principle for justified beliefs is not universally shared. So there may be reasons to resist a straightforward statistical understanding of reliability, if being reliably produced is close to sufficient for justification.

I should also note that reasons for rejecting transmissibility of justification over multi-premise inferences do not carry over to the case of single-premise inferences. So one might still worry that this sort of transmissibility gives rise to a problem, in that it requires skeptical scenarios to be justifiably rejected if ordinary perceptual beliefs are justifiably embraced. Here, I will only express sympathy for Leplin's approach of avoiding attempts to secure justification of ordinary beliefs without having to reject skeptical scenarios. It has always seemed to me that if we are unjustified in rejecting the skeptical scenario, we have already lost the most interesting part of the game.

The second question I'd like to raise concerns whether the requirement that reliable methods be used intentionally can provide a solution to reliabilism's generality problem.

It certainly seems to help in Leplin's cookbook case. Suppose that Sam forms the belief that asparagus should be peeled before cooking, on the basis of trusting a tattered white copy of

Mastering the Art of French Cooking. This of course instantiates many different method-types:

- a. trusting a book
- b. trusting a cook book
- c. trusting a tattered white cook book
- d. trusting Mastering the Art of French Cooking
- e. trusting a copy of a Mastering the Art of French Cooking that Sam bought at a yard sale

What method has Sam intentionally used here? Leplin's test for intentional use of a method is independent of whether Sam is conscious of using the method; rather, it involves determining what he would do under changed conditions. And it seems that this test does show us how to narrow the range of options considerably. We can see that a - c will be eliminated on the grounds that Sam would not, in general, trust any old books, cook books, or tattered white cook books. And the method will not be narrowed down to e, since Sam would trust other copies the book. This is good--a certain kind of psychological robustness or reality is part of what seems needed to solve the generality problem. And seeing what a person would do in alternative circumstances can certainly help us discover what method is doing the psychological work.

But in other cases, it is less clear how the "what would the agent do in other circumstances" test can help. For example, I

often form beliefs about who is in the room on the basis of recognizing people's faces. On some occasions, I form the belief that my wife is in the room. On other occasions, I form the belief that some particular one of my students is in the room. But I'm not nearly as reliable when identifying my students-- particularly toward the beginning of the semester. Sometimes I make mistakes. This raises an instance of the generality problem. When I identify my wife, am I using the same method, or a different one, from when I identify a student? And more generality questions arise along the same lines. I've noticed that I'm better at identifying female students than male ones. Are these separate methods? Does the method I use in identifying some particular student change as the semester progresses?

It seems clear that answering these questions will in some cases be crucial to deciding whether a particular recognition-based belief is reliably produced. And some ways of typing methods would yield highly counterintuitive results. For example, if the method I use in identifying my wife is counted as the same as the one I use in identifying some fairly nondescript male student at the beginning of the semester, the justification of my belief that my wife is in the room would be jeopardized. Yet it seems to me that my belief that my wife is in the room, when I've formed it on the basis of recognizing her face, is intuitively both very well justified and reliably produced.

But it is far from obvious how the various instances in which I form a belief about who is in the room by (apparent) facial recognition should be grouped into more specific methods. This is particularly clear because there seems to be a smooth continuum of cases, with ones involving people I know extremely well at one end, and ones involving casual or new acquaintances at the other.

Moreover, it is not clear that the test of what would be done under changed circumstances will help here. I'm not sure exactly how this idea is to be filled out in the present case. The way that seems most parallel to the cookbook example would be to ask, of an occasion on which I identify my wife by facial recognition, whether, for example, I would also form beliefs by facial recognition of a longtime student, or by facial recognition of a recent student, or by facial recognition of a nondescript male recent student. Given that I am inclined to form beliefs on the basis of facial recognition in all of these sorts of cases, the "what would David do" test seems to allow us to type my method in a very wide way--say, "forming beliefs about a person's identity on the basis of apparent facial recognition." Given my sorry record in certain kinds of cases fitting this description, typing my method so widely threatens to count my wife-identification as unjustified. To me, this is counterintuitive. And an analogous problem may be developed by



considering my identifications of one and the same student at different points in the semester.

One might reply that the test is getting these cases right after all. If I form beliefs about who I'm looking at in a way that's insensitive to whether I know the person well or not, it might be argued, my basic method is like hasty generalization--I'm drawing conclusions on the basis of insufficient evidence. Or it is like the method of trusting what one hears on one's TV, when one's TV does get FOX. The fact that I get it right when my wife happens to be the person in question is due to luck, not to my forming beliefs in a reliable way.

But I think that this reply would be too quick. The fact that I'm willing to form facial-recognition-based beliefs in unreliable ways about people I don't know well certainly doesn't mean that when I form facial-recognition-based beliefs about my wife, I'm insensitive to the features of her appearance which make my identifications of her so reliable. In fact, it seems to me that the intuitively correct way to type my method--at least for the purposes of reliabilist epistemology--should distinguish between the method I use in identifying my wife, and the one I use in identifying a recent nondescript male student.

Thus it seems that there is a limit to what the "changed circumstances" test can do in answering generality questions. The basic problem the test faces in the facial recognition cases

can be generalized to cases in which it is intuitively quite clear that two different methods are involved. I could, after all, be prone both to using the method of Encyclopedia-Britannica-consultation and the method of Ouija-board-consultation. On any occasion on which I consult the Encyclopedia Britannica, I might also be willing to consult a Ouija board, and vice-versa. If we ask, on some occasion where I consult the Britannica, if I would have consulted a Ouija board had one been substituted for my volume of the Encyclopedia, the answer would be "yes." But this should not lead us to type my method as "consulting either the Encyclopedia Britannica or Ouija board." So it is clear that reliabilism requires individuating methods more finely than a simple version of the "changed conditions" test can provide.

Now I don't want to suggest that Leplin believes the this test to be sufficient to solve the generality problem all on its own. He regards it as providing evidence of what method the agent intentionally employs, and that it certainly does, as we see in the cookbook case. And I also don't want to deny that the concept of intentional employment could do more work in attacking the generality problem. After all, even if I am partial to trusting both my Ouija board and the Encyclopedia Britannica, it is unlikely that I intentionally employ a disjunctive method. Intuitively, it's much more plausible that I'm simply prone to

intentional employment of two distinct methods. In the end, how much work the notion of intentional employment can do in difficult more cases will depend on a detailed articulation of that notion. At this point, it is unclear to me whether some such account can solve the generality problem.

The third question I want to raise concerns Leplin's two components of epistemic justification. They are used to handle the problems typically posed by clairvoyance and delusion cases. There are, of course, many different ways of dealing with our intuitions in these sorts of cases. Some of these--such as Goldman's actual-world- or normal-world-indexed versions of reliabilism--would accommodate the problematic intuitions while giving reliability the central role in determining justification. Leplin's way of handling the problematic intuitions involves distinguishing between two different dimensions of epistemic justification. This has also been tried, by Goldman, in response to delusion cases: he distinguishes between reliabilist "strong justification" and non-reliabilist "weak justification," with the latter representing epistemic blamelessness, and giving the intuitively attractive results in delusion cases. But Leplin's distinction is not Goldman's.

Leplin distinguishes between (A) a belief's being justified, and (B) an agent being justified in having that belief. This

distinction is not just the familiar one between an agent's actual belief being justified, and the agent being in a position to believe a proposition justifiedly. For the latter is typically intended to abstract from whether and how the agent believes the relevant proposition, whereas Leplin's (B) notion presupposes that the agent does believe the relevant proposition.

Leplin's B-justification also differs from Goldman's weak justification, in that the latter is tied directly to blamelessness, while Leplin rejects deontological conceptions of justification. Insofar as B-justification is supposed to account for our intuitions in clairvoyance-type cases, I think Leplin is exactly right to avoid this sort of deontology. When I take the clairvoyant agent to be unjustified in his belief, it is not because I think him blameworthy. He may happen to be so constituted that whenever he begins to consider reasons for distrusting his clairvoyant deliverances, he starts to itch so badly he cannot think of anything but scratching. Or he may, due to genetic or environmental factors entirely beyond his control, simply lack the mental energy to reflect critically on his clairvoyant beliefs enough to realize that he has no good reason to trust them. In general, it seems to me that it's easy to think of cases--beginning with unfortunate folks who wear aluminum foil hats to keep the government from listening to their thoughts--in which agents deserve no blame for their epistemic

predicaments, but are nevertheless epistemically unjustified in believing as they do.

Leplin's account B--which explicates a person's being justified in believing a proposition--is the one which is designed to explain the sorts of intuitions that have been taken to cut against reliabilism. It is B-justification that is lacking in Leplin's book-based analogue of the clairvoyance case (where the agent trusts a book that happens to be reliable, though the agent has no good reason for thinking the book to be reliable). And it is B-justification that is present in Leplin's analogue to the delusion case (where the agent has good reason to trust a book which, however, happens to be unreliable).

Now B-justification is not defined directly in terms of reliability. However, its definition does make reference to the sort of justification described by A, which is straightforwardly reliability-based. So it seems to me that it is worth asking whether the B-account constitutes a reliabilist answer to the problematic cases, or whether it in effect cedes ground to non-reliabilist conception of justification.

Now I don't want to imply that Leplin claims B-justification to be a reliabilist notion. And Leplin states straightforwardly that he doesn't take reliabilism to provide a complete theory of justification. But since the overall account seeks to meet objections to reliabilism, it seems to me an interesting question

whether this particular objection can be met with an essentially reliabilist condition, or whether the objection must be met by introducing a non-reliabilist sort of justification.

Leplin says that B is not intended as an internalist sort of justification. B-justification requires good reasons. And Leplin explicitly (and, I think, quite reasonably) rejects a strong sort of internalism about good reasons--the sort according to which we lose the distinction between something being a good reason and being thought to be a good reason. Thus B-justification does not reduce to justification of this strongly internalist sort.

But I would point out that not all accounts that would commonly be thought to be internalist--and certainly not all non-reliabilist accounts--lose the distinction between being a good reason and being thought to be a good reason. Evidentialist accounts, for example, should be able to maintain the distinction. So although B's "good reasons" requirement rules out a certain strong sort of internalism, it does not necessarily constitute a reliabilist requirement.

Leplin's B-justification condition specifies that S is justified in believing P if S has good reason to believe that her belief in P is epistemically justified--and this latter epistemic justification, the sort described in A, requires that the belief be reliably produced or sustained. This suggests that B does

after all represent in some sense a distinctively reliabilist kind of justification.

My question, then, is whether B-justification provides a way of answering the intuitive objections without retreating from reliabilism. To get clearer on this question, I want to concentrate on clairvoyance-type cases, or their less fanciful analogues. Will our intuitions in these cases generally turn on the agent's lack of good reason to think her belief reliably formed? Are there cases that elicit the same sort of intuitions even though the agent does have good reason to think her belief results from a reliable method?

A natural alternative explanation for our intuitions in the clairvoyance cases is that they violate some clearly non-reliabilist requirement on justification. For example, a simple evidentialist principle might require that to be justified in believing P, one must have good reason for thinking that one's belief that P is true. So it is natural to think about B-justification by considering cases, if such exist, where the agent has good reason to think her belief results from a reliable method, but nevertheless does not have good reason to think that this belief is true.

It is not obvious at first that this sort of case should exist--after all, having reason to think that a particular belief

was reliably produced or sustained would typically constitute at least some reason for thinking that the belief was true.

But I think that there are fairly plausible cases in which agents have reason to think their beliefs reliably produced, but not to think that their beliefs are true.

Suppose that Howell is in the habit of believing stories he reads in the New York Times, and retaining those beliefs on the basis of remembering the Times stories. Let us suppose that he has good reason to believe that trusting and remembering Times stories is a reliable method of forming and sustaining beliefs. Using this method, he forms the belief that it was sunny in Palestine, West Virginia on the day that Times reporter Jayson Blair reported interviewing Jessica Lynch's family. But Howell then learns that Blair never conducted the interview, and made up the details he included in the story. It seems to me that, if Howell somehow maintains his belief on the basis of his trusting and remembering the Times story, even after finding out that Blair fabricated the details, then we may have an example of the sort required: Howell would still have good reason to think that the general method by which his belief about Palestine was produced and sustained was a reliable one (after all, Blair's dishonesty should presumably count as abnormal conditions at the Times). Nevertheless, Howell would not have (overall) good reason to think that the belief was true.



Now I do have a worry about this example, which is similar to some of what Leplin worries about in clairvoyance cases: in any ordinary case of this sort, Howell would not maintain his belief about Palestine. But let us suppose that he has an emotional attachment to the Times which prevents his knowledge about Blair from interfering with his reliance on the Times-based belief-sustaining process. Or perhaps he just fails to "put two and two together," in the same way that I might know I have a dentist appointment on Tuesday afternoon, and also know that today is Tuesday, but fail to realize that I have a dentist appointment this afternoon.

If this case description is correct, then it seems that B-justification does involve reference to reliability in an important way. B's invocation of reliability makes it differ from the evidentialist-style condition specifying that a person is justified in believing a proposition only when she has reason to think that her belief is true. It is Howell's reasons for belief about the general reliability of the Times-trusting method, and not Howell's reasons for thinking his specific belief about Palestine true, that enter into B-justification and make its verdict differ from that of the more standard evidentialist condition.

However, insofar as this is so, this sort of case does not seem to me to support a B-style reliabilist answer to the problem

that clairvoyance-type cases pose. For to my mind, at least, it seems that Howell is not intuitively justified in believing that it was sunny in Palestine on that day. And it seems to me that our reasons for denying that Howell is justified in believing it was sunny in Palestine are essentially similar to our reasons for denying justification in clairvoyance cases, or in the analogous book-based case. Suppose a person trusts a book for extraneous personal reasons, and forms a belief that P. We intuitively think him unjustified in believing P because he has no reason to think that the book told him the truth about P. And we think that Howell is unjustified in his belief about Palestine because Howell has no good reason to think that the Times story told him the truth about the weather in Palestine. The fact that his Palestine belief is known to be produced by a reliable method cuts no intuitive ice, once its typical inductive force as evidence of the belief's truth is undermined. This suggests that the reliability component of B-justification does its intuitive work only insofar as it goes proxy for a more basic evidentialist condition.

Of course, this is just one example. Not only might one reject my analysis of it, but there may well be other examples that favor B over the simple evidentialist alternative. But the example seems to me at least to raise the possibility that B-justification be amended to eliminate even the indirect reference

to reliable belief production. If this were done, of course, the two-component solution to the clairvoyance problem would be a straightforward concession to a non-reliabilist understanding of justification.

From the point of view of some reliabilists, this might be quite acceptable. Even Goldman said that he wouldn't mind if his two-component understanding of justification was a "marriage of externalism and internalism". From the point of view of one who is suspicious of reliabilism, it of course suggests the possibility of developing arguments showing that the reliabilist component of the theory was left without sufficient work to do. Fortunately for me, however, pursuing that issue is quite clearly beyond the scope of the present discussion.