

Internalism and Evidence of Reliability

Anthony Brueckner

Received: 18 July 2007 / Revised: 18 March 2008 / Accepted: 31 March 2008 /
Published online: 26 April 2008
© Springer Science + Business Media B.V. 2008

Abstract This paper concerns various competing views on the nature of perceptual justification. Various thought experiments that motivate various views are discussed. Once reliabilism is rejected and some form of internalism is instead embraced, the following issue arises: must an internalist nevertheless require that perceptual justification involve the possession of evidence for the reliability of our perceptual processes? Matthias Steup answers in the affirmative, espousing what he calls *internalist reliabilism*. Some problems are raised for this form of internalism.

Keywords PPR · Standard Demon World · RTB · Internalist reliabilism · Reverse Demon World

Pure Process Reliabilism and the Standard Demon World

What is the role of sense experience in the justification of perceptual beliefs—beliefs that are based upon experience? On a *pure process reliabilist* (PPR) approach to justification in general, $R=J$, where R is the condition that one's belief issues from a sufficiently reliable belief-forming process, and J is the condition that one's belief is justified.¹ On PPR, $K=RTB$ (knowledge is reliably produced true belief). So on PPR, my perceptual belief that my cat is on my lap is justified in virtue of the fact that it issues from the highly reliable *visual* belief-forming process. In a well-known discussion, Stewart Cohen pointed to a very useful way of testing our intuitions about externalist theories of justification such as PPR.² Suppose that I am in a normal world, and, as I believe, my perceptual beliefs issue from highly reliable belief-forming processes. Now consider my mental twin who inhabits what we will

¹See, e.g., A. Goldman, "What is Justified Belief?", in G. S. Pappas (ed.), *Justification and Knowledge: New Studies in Epistemology* (Dordrecht: Reidel, 1979).

²See S. Cohen, "Justification and Truth", *Philosophical Studies* 46 (1984), 279–95.

A. Brueckner (✉)

Department of Philosophy, UCSB, Santa Barbara, CA 93117, USA
e-mail: brueckne@philosophy.ucsb.edu

call the *Standard Demon World*, in which the twin's perceptual beliefs issue from belief-forming processes that are completely *unreliable*. My twin has experiences that are qualitatively indistinguishable from mine, beliefs, thoughts, desires, and intentions that are just like mine, and so on.³ In evaluating the demon-world *truth-ratios* of my twin's perceptual belief-forming processes, the numerator of the pertinent ratio is always set at *zero*. A powerful *internalist* intuition is elicited in many philosophers by consideration of this pair of worlds. As Cohen says,

[In the demon world] our experience is just as it would be if our cognitive reasoning processes were reliable, ...[In the demon world] we would have every reason for holding our beliefs...that we have in the actual world. Moreover since we actually have reason to believe that our cognitive processes are reliable, it follows that in the demon world we would have every reason to believe that our cognitive processes were in fact reliable.⁴

Cohen's intuition is that in light of the foregoing similarities between me in the normal world and my twin in the demon world, and in light of the apparent fact that my perceptual beliefs are justified in the normal world, the conclusion to draw is that my twin's perceptual beliefs in the demon world are also justified. If this is correct, then PPR is refuted: on this view my normal-world beliefs are justified in virtue of their reliability, while my demon-world beliefs are *unjustified* in virtue of their unreliability. In what follows, we will consider a number of cases that motivate refinements of the internalist approach to the concept of justification.⁵ Just as in other areas of philosophy, we will follow the lead of our intuitive judgments about cases in analyzing the concept of justification.

R without J: the Reverse Demon World

If Cohen's intuition is on the mark, then we have $\sim(R=J)$, since in the demon world, we have $(J \text{ and } \sim R)$. Cohen's thought experiment is the starting point for Matthias Steup's very fruitful discussion of various approaches to perceptual justification.⁶ Steup is an internalist, and he shares Cohen's intuition about the Standard Demon World thought experiment. According to Steup, not only do we have $\sim(\text{If } J, \text{ then } R)$, but, further, we have $\sim(\text{If } R, \text{ then } J)$. One way of arguing the latter is BonJour's.⁷

³ I will ignore Putnamian worries about whether such a twin is really possible: one might hold that the difference in environments between me and my "twin" would induce differences in the contents of our mental states, so that my "twin" does not, e.g., have the same beliefs as me.

⁴ See "Justification and Truth", 281.

⁵ It has been suggested that there are two distinct epistemic concepts that are lumped under the heading of 'justification': an internalist concept that applies to mature humans, and a broader externalist concept that applies to children and higher non-human animals as well as to human sophisticates. It may be that the investigations of this paper, then, concern just the internalist concept.

⁶ See M. Steup, "Internalist Reliabilism", *Philosophical Issues* 14 (2004), 403–25.

⁷ See L. BonJour, "Externalist Theories of Empirical Knowledge", in P. French, T. E. Uehling, and H. Wettstein (eds.), *Midwest Studies in Philosophy V* (1980), 53–73. Goldman refined the basic reliabilist analysis he put forward in "What Is Justified Belief?" in his book *Epistemology and Cognition* (Cambridge: Harvard University Press, 1986). One change he made was designed to handle the BonJour clairvoyant example.

We consider a case in which, say, Madonna has a perfectly reliable clairvoyant power regarding the whereabouts of the President. Madonna never bothers to check up on the deliverances of the little voice in her head—she is simply too busy. One day, her little voice whispers that NY=The President is in New York City. Madonna has no reason to believe NY, and, we will suppose, she has a mountain of good evidence against NY. The news media, along with Madonna's White House contacts, proclaim that the President is in Washington, D.C., even though Jack Bauer has in fact spirited him away to New York City to protect him from a terrorist threat. Still, Madonna goes with the little voice (whose deliverances she has never investigated) and so believes NY. According to PPR, Madonna has a JTB of NY because she has an RTB of NY. So on PPR, Madonna knows NY. Many philosophers, however, have the firm intuition that Madonna's belief of NY is not justified since it is irrationally held in the face of strong counter-evidence and, further, lacks any positive evidential basis. So we have (R and \sim J), and Madonna does not know NY.

Steup's alternative way of making a case against the sufficiency of R for J will take us into the issues that are the main focus of the present paper. Steup has us consider the Reverse Demon World, in which our belief-forming processes are highly reliable but in which a Cartesian demon gives us compelling evidence for believing that our perceptual faculties are unreliable.⁸ In such a world, according to Steup, any perceptual belief formed on the basis of an in-fact-reliable perceptual process would lack justification due to the (misleading) evidence of unreliability that the demon foists upon us.⁹ The Reverse Demon World is structurally similar to the Madonna example. In such a world, Madonna's situation holds across the board: in-fact-reliability paired with misleading counter-evidence.

Undefeated

So for Steup, reliability is not required for perceptual justification (see Standard Evil Demon World), and reliability is not sufficient for perceptual justification (see Reverse Evil Demon World). A necessary condition for perceptual justification that emerges from consideration of the Reverse Evil Demon world, according to Steup, is this: perceptual justification requires the absence of evidence of global perceptual unreliability, which evidence I would have in the Reverse Demon World. What would evidence of global perceptual unreliability be like? Suppose that right now I seem to see a cat on my desk and then attempt to feel the cat's fur. If my hand seems to pass through the cat, then this would be evidence that my visual perceptual faculty is not functioning well on this occasion—evidence of local unreliability. Similarly, if I blink my eyes and then cannot see any cat in my closed study, then this would also be evidence of current malfunctioning of my vision—again, evidence of local

⁸ We will discuss what such evidence would look like shortly.

⁹ One possibility is to hold that our perceptual beliefs in this demon world have *prima facie* justification, conferred upon them by their reliability, which justification is defeated by the evidence of unreliability. This view departs from PPR, according to which reliability is sufficient for all-things-considered justification.

unreliability.¹⁰ But more is required for me to have evidence that my perceptual faculties are globally unreliable in the sense that Steup has in mind. I would need to have memories of a dismal track record of perceptual failures of the sorts I just now experienced regarding the cat. Alternatively, if I have memories of completely chaotic past experiences, then this could count as evidence of global perceptual unreliability.¹¹ The important point for what will emerge later is that Steup emphasizes the role of memory in evidence regarding the global reliability of my perceptual faculties (or lack thereof).¹²

We can now refine the necessary condition on perceptual justification that emerges from consideration of the Reverse Demon world: one must lack memories of perceptual unreliability. But this is not quite right, as Steup points out. In the Reverse Evil Demon world, I do not have memories of perceptual unreliability, since my perceptual faculties are in fact reliable. So what we must instead require is that I lack memory impressions (which might not be memories) of global perceptual unreliability. Call this condition on perceptual justification *Undefeated*.

Certified

Steup wants to put forward a further necessary condition on perceptual justification. I will first describe a case which Steup does not discuss. Suppose that on my fortieth birthday, I have a series of vivid experiences as of seeing a cat on my desk and then petting the cat. I form the appropriate perceptual beliefs on the basis of these experiences. Are these beliefs justified? We know that for Steup, the question whether the beliefs are produced by perceptual belief-forming process that are in fact reliable is irrelevant to our question of justification. Suppose I attempt to consult my memory regarding past perceptual experiences and beliefs. I have been seized with amnesia and cannot dredge up any memory impressions at all regarding the past reliability of my perceptual faculties (or lack thereof). So I satisfy the condition *Undefeated* for having perceptual justification. But for Steup, another condition must also be satisfied. Steup thinks that in addition to satisfying the purely negative *Undefeated*, I must satisfy a further positive condition: I must have evidence for the reliability of my perceptual faculties. Call this condition *Certified*. So in the amnesia case, my cat-beliefs are not justified, since I fail to satisfy *Certified*. The more precise condition Steup has in mind will again treat of memory. As a first shot, we will hold that perceptual justification requires that one have memory impressions of a track record of perceptual success. Note that (a) the memory impressions might be

¹⁰ Evidence of such local unreliability would preclude my cat-beliefs in the foregoing cases from having justification.

¹¹ Of course, if I inhabit what is in fact a completely chaotic world, then the right kind of chaotic experience will issue from reliable perceptual processes, and the pertinent beliefs will be true beliefs. Think of the experience of a human space traveler such as that depicted at the end of Stanley Kubrick's *2001: A Space Odyssey*.

¹² An anonymous referee raised the question whether a mental state having complete chaos as its content could count as a *memory impression*. Following Kant in the *Refutation of Idealism*, we might well inquire into the conditions that are required for a mental state to represent a past series of temporally ordered events. Such conditions might involve representation of stable, causally ordered sequences of events.

unveridical, and (b) even if the memory impressions are veridical, it could still be that my perceptual faculties are in fact completely unreliable. What goes for my perceptual faculties also goes for my memory: in addition to having evidence of perceptual reliability, perceptual justification also requires having evidence of the reliability of the deliverances of memory which seem to certify the reliability of perception. We can now state Steup's Internalist Reliabilism:

INREL. A subject's sense experiences are a source of justification for her iff her memory impressions collectively depict a track record of both perceptual and memorial success.¹³

Note again that the memory impressions of memorial success might be unveridical, as well as the memory impressions of perceptual success.

For Steup, the only contenders for a reasonable account of perceptual justification are INREL and the more minimal account that merely requires Undeclared without requiring Certified as well. Steup discusses a number of cases which seem to him to favor INREL over the more minimal account. I will briefly discuss just one and then raise a worry about INREL.

Suppose that a neurosurgeon is going to attempt to equip me with a perceptual faculty of echolocation. I am told that the operation may work, and it may fail, in which case I will have some quasi-visual, novel unveridical experiences which will not derive from a new echolocating faculty. After the operation, I have a quasi-visual, novel experience as of a table in the room. Steup's intuition is that absent any evidence that goes towards certifying the reliability of whatever perceptual faculty is at work, I am not justified in believing that there is a table in the room. This is so even though I satisfy Undeclared with respect to whatever faculty is at work.

Internalist Intuitions

Let us take stock. Steup's internalist intuitions lead him to maintain that my mental twin in the Standard Demon World is justified in holding his perceptual beliefs despite the unreliability of his perceptual belief-forming processes. Reliability is not necessary for perceptual justification. Steup maintains that in the Reverse Demon World, in which perceivers have misleading evidence of perceptual unreliability, these in-fact-reliable perceivers are unjustified in holding their perceptual beliefs. Reliability is not sufficient for perceptual justification. Steup maintains that a perceiver who lacks memory impressions which certify an apparent track record of perceptual reliability (such as amnesiacs and wanna-be echolocators) is unjustified in holding his perceptual beliefs.

What is driving these intuitive judgements in the three sorts of case? In the Standard Demon World, my mental twin is internally just like me: I am (we will assume) a subject with a normal mental and perceptual life, in which I seem to see and do see cats, cars, clotheslines, and so on. In the Reverse Demon World, perceivers are not internally just like me (and other normal perceivers), since they

¹³ Note that the memory impressions of memorial success might be unveridical, as well as the memory impressions of perceptual success.

have evidence of perceptual unreliability, unlike me and other normal perceivers. In amnesia and wanna-be echolator cases, perceivers again are not internally just like me (and other normal perceivers), since these subjects lack normal certifying memory impressions of perceptual reliability. It seems that Steup's internalist intuitions about these three sorts of case begin with, first, the assumption that normal perceivers such as me are by and large justified in holding our perceptual beliefs. Second, Steup's intuition is that subjects who are internally like such normal perceivers are justified in holding their perceptual beliefs, regardless of their external circumstances (such as the unreliability of their perceptual belief-forming processes). Subjects who are internally different from such normal perceivers in pertinent respects are unjustified in holding their perceptual beliefs, regardless of their external circumstances (such as the reliability of their perceptual belief-forming processes). So there look to be two sources from which Steup's intuitive judgements flow.

We have seen that according to another aspect of Steup's view, the memory impressions that "certify" the past reliability of a justified subject's perceptual belief-forming processes need not be veridical memory impressions. The "certifying" memory impressions must attest to the perceiver's past reliability, but they need not do so in a veridical manner, on Steup's view. This is in keeping with Steup's anti-externalist intuitions. Just as a justified perceiver's experiences need not issue from in-fact-reliable belief-forming processes, his "certifying" memory impressions need not be reliable indicators of the character of his past experiences. These memory impressions can be thoroughly unveridical while nevertheless doing their "certifying" work.

A Problem for INREL: the Really Nasty Demon World

Suppose that I have 40 years worth of wildly chaotic experience, or maybe 40 years of experiences like those involving the untouchable, and later disappearing, cat.¹⁴ So let my experiences be just as represented by my unveridical memory impressions in the Reverse Demon World—experiences which provide evidence of the unreliability of my perceptual faculties. Suppose further that during this long stretch of time, I am besieged with unveridical memory impressions which misrepresent my run of experiences. My life is simply a cognitive and experiential mess. On my fortieth birthday, I have a two minute series of coherent experiences as of seeing waves breaking on the beach. (Maybe I am suddenly on a beach, maybe not). Suppose that during this brief stretch of time on my birthday, I consult my memory and find a rich store of unveridical memory impressions which present a sterling record perceptual and memorial success. My mental life is the product of a really nasty demon who wants me to mistakenly think, on my birthday for the two minute stretch of beach-experience, that I have had a normal perceptual and memorial life. Nothing could be further from the truth! After my two minutes of coherent beach-experience, my really nasty demon flips the haywire switch, and my experiential and memorial disaster resumes for the duration of my life.

In this case, I satisfy Steup's INREL during the two minute series of beachish experiences. But it is not at all clear that my birthday beliefs about waves breaking

¹⁴ On Steup's view, it is irrelevant whether my perceptual faculties in this imagined world are in fact reliable (due to my inhabiting a crazy world).

on the beach are justified. One strand in Steup's internalism supports the intuitive judgement that my birthday beliefs lacks justification. In the case at hand, I am very different from a normal perceiver, since my experiences in general provide no evidence of perceptual reliability (except for the two minutes of beach experience), and my certifying memory impressions of perceptual and memorial reliability are all unveridical. Suppose that the sheer weirdness of my situation in the case at hand successfully grounds the intuitive judgement that my birthday beliefs are unjustified and that INREL is thus mistaken, even by broadly internalist lights. Then an "internalist reliabilist" along the lines of Steup must somehow take into account the history of a subject's perceptual experiences. *In the Really Nasty Demon World*, my experiences unfold chaotically through time in such a way as to prevent my beach-beliefs from possessing justification, even though I satisfy INREL at the pertinent times (during the two minute stretch of beach-experience). But how we are to take history into account is unclear. Do we simply stipulate that perceptual justification requires that (1) the subject satisfy INREL's Certified condition, and (2) the subject in fact have normal, orderly experience like mine and yours?

(2) Is not only vague, but, further, it engenders implausible consequences such as the following. Imagine a world populated by a great many holograms, as well as by a great many ordinary objects. Imagine that perceivers learn that all the things that look like our non-human animals are holograms through which the perceivers can pass their hands and which "blink" on and off. The experiences of the perceivers in the Hologram World are very different from ours, and thus (2) is violated. So the perceivers in the Hologram World lack perceptual justification, according to the revised account. But this seems very implausible.

Perhaps Steup should simply accept INREL's implication: my birthday beliefs are justified. My decidedly abnormal perceptual and memorial history in the Really Nasty Demon World is just not relevant to the question whether my birthday beliefs are justified. All that is required in order for those beliefs to be justified is that the beliefs present the appearance of issuing from historically reliable perceptual belief-forming processes. The birthday beliefs in fact issue from unreliable processes—no problem. My perceptual processes are "certified" by memory impressions which are in fact unveridical and in fact issue from unreliable memorial processes—no problem. My experiential and cognitive life is in fact a blooming, buzzing confusion—no problem. For one brief, shining moment, I am blessed with justified perceptual beliefs. It may be that the sad history of my cognitive life is irrelevant to the question whether I have justification for my birthday beliefs: to answer that question, we confine ourselves to a scrutiny of my internal states during the period of time when I hold those beliefs.

If the case of the birthday beliefs nevertheless does seem to constitute a counterexample to Steup's INREL, one possibility for the internalist is to modify INREL in the following way:

INREL*. A subject's sense experiences are a source of justification for her iff her memory impressions are veridical and collectively depict a track record of both perceptual and memorial success.¹⁵

¹⁵ Thanks to an anonymous referee for this suggestion.

INREL* requires that the experience of a justified subject S resemble our orderly, normal experience only in the respect that S's experience must provide evidence for the reliability of his perceptual belief-forming processes. This requirement is induced by INREL*'s requirement that S have veridical memory impressions of a track record of perceptual success—a track record that is represented by my unfolding experiences. Our justified S could well be in a Standard Demon World, for all that INREL* requires: S's experiences need not issue from in-fact-reliable perceptual belief-forming processes. They just must seem to, and this appearance must be later captured by veridical memory impressions of S.

If Steup were to opt for INREL*, then his internalism would include an externalist element. This is because INREL* combines internalist anti-reliabilism about justification-conferring perceptual processes with a form of reliabilism regarding certifying memory impressions.

So where does this leaves us? The thoroughgoing internalism of Steup's INREL has the unattractive consequence that my birthday beliefs are said to be justified. The alternative INREL* avoids that consequence at the price of watering down thoroughgoing internalism, yielding a sort of gerrymandered internalist/externalist position.¹⁶

¹⁶ Thanks to Stewart Cohen, Craig Evans, and three anonymous referees for helpful comments.