How I Know I'm Not a Brain in a Vat*

JOSÉ L. ZALABARDO

1. Introduction

The problem of scepticism, as it figures in contemporary epistemology, takes the form of a series of arguments for the conclusion that we don't have much of the knowledge that we think we have. Some of the most prominent arguments for this conclusion take as their starting point *sceptical hypotheses*. Perhaps the most famous of these is Descartes' evil demon hypothesis, according to which

> [. . .] some malicious demon of the utmost power and cunning has employed all his energies in order to deceive me. [. . .] the sky, the air, the earth, colours, shapes, sounds and all external things are merely delusions of dreams that he has devised to ensnare my judgement.[1]

Hilary Putnam's brain-in-a-vat hypothesis (BIV) offers a contemporary variation on the Cartesian theme:

> [. . .] imagine that a human being (you can imagine this to be yourself) has been subjected to an operation by an evil scientist. The person's brain (your brain) has been removed from the body and placed in a vat of nutrients which keeps the brain alive. The nerve endings have been connected to a super-scientific computer which causes the person whose brain it is to have the illusion that everything is perfectly normal. There seem to be people, objects, the sky, etc; but really all the person (you) is experiencing is the result of electronic impulses travelling from the computer to the nerve endings.[2]

* I have presented versions of this paper at the Arché Centre of the University of St Andrews, the Cambridge Moral Sciences Club, the Royal Institute of Philosophy 2006–7 lecture series and the conference organised by the Instituto de Investigaciones Filosóficas in Mexico to celebrate the 40th anniversary of the journal *Crítica*. I am grateful to these audiences, especially to Mike Martin. Support from the British Academy to attend the *Crítica* conference is gratefully acknowledged.

[1] R. Descartes 1984: 15.

[2] H. Putnam 1981: 5–6.

doi:10.1017/S1358246109000071 © The Royal Institute of Philosophy and the contributors 2009

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

There are several ways in which sceptical hypotheses can be used in an argument for the sceptical conclusion that I have very little knowledge. One that has received a good deal of attention in recent debates seeks to draw a conclusion to this effect from two thoughts concerning sceptical hypotheses, namely, that we don't know that they don't obtain and that if we don't know this then there is a lot else that we don't know either.[3] I shall refer to this line of reasoning as the *canonical sceptical argument*.[4]

We can present the structure of the argument in a typical instance using the brain-in-a-vat hypothesis:

P1: I don't know not-BIV.
P2: If I don't know not-BIV, I don't know that I have a broken fingernail.

Therefore:

C: I don't know that I have a broken fingernail.

It is unquestionable that the premises of the canonical argument entail its conclusion. Hence resisting the conclusion would seem to require rejecting at least one of the premises. Both strategies enjoy support.

Rejecting the second premise is the strategy for resisting the canonical argument endorsed by Robert Nozick and Fred Dretske.[5] One serious obstacle to their approach is the fact that in many instances of the canonical argument the second premise follows from the principle that knowledge is closed under known entailment:

Closure: If S knows that p and S knows that p entails q, then S knows that q.

Clearly P2 follows from Closure, since I know that my having a broken fingernail entails not-BIV. Hence pursuing the Dretske/Nozick approach to the canonical argument requires rejecting Closure.[6] Both authors have argued that this move can be independently motivated, but many contemporary epistemologists disagree.

[3] For an alternative approach to the use of sceptical hypotheses in sceptical arguments see A. Brueckner 1994.

[4] Cf. Ibid., 827.

[5] Cf. F. Dretske 1970, R. Nozick 1981: Ch. 3.

[6] As DeRose has observed, in some instances of the canonical argument the second premise seems compelling even though it doesn't follow from Closure – because the proposition that figures in the consequent is compatible with the sceptical hypothesis. See K. DeRose 1995: 32 fn. See also J. Pryor 2000: 522.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

On the contrary, they regard the rejection of Closure as too high a price to pay for victory over the sceptic – at least if other strategies are available.

The alternative strategy is to reject the first premise of the canonical argument, arguing in each case that we know that the sceptical hypotheses don't obtain, or, at least, that the sceptic hasn't shown that we don't. My main goal in this paper is to defend a version of this line of thought.

2. DeRose and the Canonical Argument

I want to introduce my proposal by looking at the strategy for dealing with the canonical sceptical argument advanced by Keith DeRose. DeRose is of course best known for his contextualist account of the semantics of sentences ascribing or denying knowledge – the view that their content varies with the context in which they are uttered.[7] But in his treatment of the canonical argument he deploys several interesting ideas that don't depend on his contextualist views. The strategy that I am going to put forward will be similar in important respects to the position that would result if DeRose's ideas were purged of their contextualist aspects.

Let me start by introducing some central features of DeRose's position. As I've just mentioned, DeRose believes that the content of sentences ascribing or denying knowledge depends on the context in which they are uttered. The way in which context affects the content of these sentences is, according to DeRose, by determining how strong an epistemic position you need to be in in order to make a knowledge-ascribing sentence true. While in everyday contexts the standards are relatively undemanding, in contexts in which sceptical hypotheses are under discussion they are much more exacting. Here is DeRose's explanation of the crucial notion of the strength of one's epistemic position, in terms of the possible-world idiom:

> An important component of being in a strong epistemic position with respect to P is to have one's belief as to whether P is true match the fact of the matter as to whether P is true, not only in the actual world, but also at the worlds sufficiently close to the actual world. That is, one's belief should not only be true, but should be non-accidentally true, where this requires one's belief as to whether P is true to match the fact of the matter at

[7] Cf. K. DeRose 1995, K. DeRose 1999.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

> nearby worlds. The further away one can get from the actual world, while still having it be the case that one's belief matches the fact at worlds that far away and closer, the stronger the position one is in with respect to P.[8]

Knowledge, according to DeRose, requires a certain level of strength in your epistemic position with respect to the known proposition. Thus, in terms of DeRose's construal of the strength of one's epistemic position, it requires that your belief tracks the truth in a certain sphere of possible worlds, centred in the actual world, to which he refers as the *sphere of epistemically relevant worlds*.[9]

Now we can provide a precise characterisation of how context affects the content of sentences ascribing or denying knowledge on DeRose's position. Context, according to DeRose, determines how far the sphere of epistemically relevant words extends – how far into counterfactual space your belief needs to track the truth in order to count as knowledge.

In everyday contexts, the sphere of epistemically relevant words doesn't extend very far. In order to make a knowledge-ascribing sentence true, when uttered in such a context, your belief would only have to track the truth in a relatively reduced range of worlds. And both my belief that I have a broken fingernail and my belief that I am not a brain in a vat satisfy this requirement. This fact provides the key to DeRose's rejection of the challenge to our everyday knowledge claims posed by the canonical argument.

DeRose's semantics for the verb 'to know' get in the way of a straightforward presentation of his anti-sceptical strategy.[10] In order to overcome this difficulty, I propose to introduce a context-insensitive neologism to express how, according to DeRose, a subject S has to be related to a proposition p in order to make 'S knows that p' true when uttered in an everyday context. Thus, let 'DR(E)-knows' be a binary predicate that is true of a subject S and a proposition p just in case S's belief that p tracks the truth of p in the sphere of epistemically relevant worlds in force in everyday contexts.

Now, in order to challenge the claim that I would express in an everyday context with the sentence 'I know that I have a broken fingernail', the sceptic would have to establish the following conclusion:

CE: JZ doesn't DR(E)-know that he has a broken fingernail.

[8] K. DeRose 1995: 34.
[9] Ibid. 37.
[10] See, in this connection, Ibid. 40 fn.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

And to establish this conclusion with the canonical sceptical argument, she would have to invoke the following premises:

P1E: JZ doesn't DR(E)-know not-BIV.
P2E: If JZ doesn't DR(E)-know not-BIV, then he doesn't DR(E)-know that he has a broken fingernail.

DeRose's treatment of this argument is an instance of the strategy that I want to recommend. He feels entitled to resist the conclusion of this valid argument because, even though he believes the second premise to be true, he thinks that the first premise is false.[11] In the sphere of epistemically relevant worlds in force in everyday contexts, my belief in not-BIV does track the truth.

Once we have defused the argument for CE in this way, we have removed the pressure that the canonical argument placed on our everyday knowledge claims. And with this, DeRose thinks, the main job of refuting the sceptic can be brought to an end. For the reason why sceptical arguments pose a serious problem is that they threaten to show,

> [...] not only that we fail to meet very high requirements for knowledge of interest only to misguided philosophers seeking absolute certainty, but that we don't meet even the truth conditions of ordinary, out-on-the-street knowledge attributions. They thus threaten to establish the startling result that we never, or almost never, truthfully ascribe knowledge to ourselves or to other mere mortals.[12]

I want to highlight the fact that in this defence of our everyday claims to knowledge from the threat of the canonical argument no role is played by DeRose's contextualism. The strategy would still be available to someone who thought that the sphere of epistemically relevant worlds is fixed for every context – that an utterance of 'S knows that p', in any context, is true just in case S DR(E)-knows that p.[13]

Nevertheless, contextualism still has an important role to play in DeRose's overall treatment of the canonical argument. For the goal of his strategy is twofold: "To safeguard ordinary claims to know while at the same time explaining the persuasiveness of the skeptical

[11] "Thus, on our solution, we do know, for instance, that we are not BIVs, according to ordinary standards of knowledge" (Ibid. 39).

[12] Ibid. 4. See also his remarks on the timid sceptic on 5–6.

[13] In a recent paper, DeRose has spelt out the connections between his position and non-contextualist views that reject the first premise of the canonical argument. See K. DeRose 2004.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

arguments [...]".[14] As we have seen, contextualism has no role to play in the first of these tasks. It is in the second task – explaining the persuasiveness of sceptical arguments – that DeRose's strategy makes use of contextualism. Given that he is proposing to resist the canonical argument by rejecting the first premise, naturally his main challenge in discharging the second task is to explain what makes this premise so appealing, "to explain what it is about sceptical hypotheses that makes it so plausible to suppose that we don't know that they're false".[15]

The crucial point here is that on DeRose's account, when the sceptic utters 'JZ doesn't know not-BIV', she immediately generates a conversational context in which the extent of the sphere of epistemically relevant worlds is radically expanded – to include worlds in which BIV is true.[16] But in those worlds I believe not-BIV – it is a salient feature of sceptical hypotheses that if they were true I would still believe them to be false. Hence, my belief in not-BIV doesn't track the truth in the sphere of epistemically relevant worlds that is in force in this context. Therefore, when the sceptic utters 'JZ doesn't know not-BIV' she is speaking the truth. Furthermore, my utterance of 'I know not-BIV' would have the same effect on the conversational context. This is DeRose's explanation of why we find the first premise of the canonical argument so plausible:

> [...] we are able to explain its plausibility [...] by means of the fact that the high standards at which (1) [the first premise of the canonical argument] is true are precisely the standards that an assertion or denial of it put into play. Since attempts to assert (1) are bound to result in truth, and attempts to deny it are destined to produce falsehood, it is no surprise that we find it so plausible.[17]

Notice that the truth that is expressed by an utterance of 'JZ doesn't know not-BIV' is not the proposition that would enable the sceptic to establish CE with the canonical argument. Let me introduce one more binary predicate, 'DR(H)-knows', true of a subject S and a proposition p just in case S's belief that p tracks the truth of p in the sphere of epistemically relevant worlds in force when

[14] K. DeRose 1995: 6.
[15] Ibid. 17.
[16] DeRose's ingenious account of how this expansion of the sphere of epistemically relevant worlds comes about is his Rule of Sensitivity. See Ibid. 36.
[17] Ibid. 40.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

my knowledge of not-BIV is asserted or denied. Then, the sceptic's utterance of 'JZ doesn't know not-BIV' expresses the proposition:

P1H: JZ doesn't DR(H)-know not-BIV.

To get from here to CE, the sceptic would need, as an additional premise, the proposition:

P2HE: If JZ doesn't DR(H)-know not-BIV, then he doesn't DR(E)-know that he has a broken fingernail.

But P2HE is false. Even though my belief in not-BIV doesn't track the truth in the sphere of epistemically relevant worlds instituted by discussion of my knowledge of not-BIV, my belief that I have a broken fingernail tracks the truth in the sphere of epistemically relevant worlds in force in everyday contexts.

The (true) proposition that would be expressed by uttering in the sceptical context the sentence 'If JZ doesn't know not-BIV, then he doesn't know that he has a broken fingernail' is rather:

P2H: If JZ doesn't DR(H)-know not-BIV, then he doesn't DR(H)-know that he has a broken fingernail.

From this, the sceptic can derive

CH: JZ doesn't DR(H)-know that he has a broken fingernail.

This is the true proposition that would be expressed by an utterance in the sceptical context of the sentence 'JZ doesn't know that he has a broken fingernail'. Hence, when the sceptic asserts her conclusion, she speaks the truth, but not the putative truth (CE) that she was aiming to establish, but the much less disturbing truth that, with respect to the state of my fingernail, I fail to meet "very high requirements for knowledge of interest only to misguided philosophers seeking absolute certainty". In sum, DeRose's explanation of the plausibility of the canonical argument doesn't reinstate the threat to our everyday knowledge claims.

Now, I am what DeRose calls a nonsceptical invariantist.[18] I think that utterances of 'S knows that p' express the same proposition in all contexts, and that this proposition is quite close to the proposition that DeRose thinks they express in everyday contexts. This means that I'll be able to join DeRose in vindicating our everyday knowledge claims in the face of the challenge of the canonical argument by rejecting its first premise. I know not-BIV, or at least the sceptic hasn't shown that I don't. My goal in the remainder of this paper is to

[18] Cf. K. DeRose 1999: 192.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

spell out the features of the concept of knowledge that enable us to adopt this position.

The price I have to pay for rejecting DeRose's contextualist semantics for the verb 'to know' is that I won't be able to avail myself of his explanation of the intuitive plausibility of the canonical argument. I face the challenge that DeRose poses for straightforward (i.e. non-contextualist) solutions to the problem:

> To succeed, a straightforward solution must explain what leads our intuitions astray with respect to the unlucky member of the triad [the premises of the canonical argument and the negation of its conclusion] which that solution denies.[19]

In my case, the challenge consists in explaining our intuitive reluctance to claim what I hold to be true – that we know that sceptical hypotheses don't obtain. This is a challenge that I accept, but the task of meeting it will be left for another occasion.

3. The Risk of Error

A very interesting aspect of DeRose's position is a contrast between the way I DR(E)-know that I have a broken fingernail and the way I DR(E)-know not-BIV. I DR(E)-know the former by virtue of the fact that my belief that I have a broken fingernail is *sensitive* to the truth of the proposition that I have a broken fingernail – i.e. if I didn't have a broken fingernail I wouldn't believe that I do. By contrast, for DR(E)-knowledge of not-BIV, sensitivity is not needed. I DR(E)-know this by virtue of the fact that the sphere of epistemically relevant worlds contains only worlds in which not-BIV is true and I believe it. My belief is not sensitive, but the nearest worlds in which its insensitivity is manifested (BIV worlds) lie outside the sphere of epistemically relevant worlds.

We have then that, on DeRose's position, what it takes to DR(E)-know a proposition depends on whether or not the sphere of epistemically relevant worlds contains worlds in which the proposition is false. If, on the one hand, there are no such worlds, it would be enough for you to believe the proposition in every world in this sphere. If, on the other hand, there are such worlds, your belief will need to be sensitive.

I think that this comes very close to adequately grasping an important intuition about knowledge – that *a true belief won't have the status*

[19] K. DeRose 1995: 42.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

of knowledge if there is a substantial and uncontrolled risk of the belief being in error. I shall refer to this as the *Risk of Error* (*ROE*) *constraint*.

The first feature of the ROE constraint that I want to highlight is the fact that what it takes to satisfy it depends on whether a substantial risk of error exists. If, on the one hand, there is no substantial risk of a belief being in error, the constraint is immediately satisfied. This corresponds, in DeRose's account of knowledge in everyday contexts, to the fact that when a belief is true throughout the sphere of epistemically relevant worlds, it doesn't need to be sensitive in order to have the status of knowledge. If, on the other hand, a substantial risk of error exists, satisfying the constraint requires bringing the risk under control. This is achieved, in DeRose's picture, when the belief is sensitive. My belief that not-BIV satisfies the ROE constraint in the first of these ways, and my belief that I have a broken fingernail in the second. In this section I want to offer a construal of the ROE constraint that will be similar in outline to DeRose's account of the strength of one's epistemic position required for knowledge in everyday contexts, but will depart from DeRose's position in a few important respects.

Let me start by fixing the sphere of epistemically relevant worlds for all contexts roughly at the level at which DeRose would place it for everyday contexts. I shall refer to it as the *Relevant Sphere*. It contains worlds in which my fingernail is not broken, worlds in which I have no hands, etc. but not worlds in which I am a brain in a vat or a victim of a Cartesian evil demon. Notice, though, that the Relevant Sphere doesn't exclude demon and BIV worlds by definition. It excludes them only if they are indeed as distant from the actual world as we think they are. Clearly, this description of the Relevant Sphere leaves a huge scope for borderline cases, and this feature will be inherited by the constraint on knowledge that I am going to formulate in terms of it.

Whether your belief satisfies the ROE constraint will depend exclusively on what happens within the Relevant Sphere. This reflects the fact that satisfying the constraint requires bringing the risk of error under control only when it is substantial. According to this approach, in order to determine whether a belief satisfies the ROE constraint, the first factor that we'll need to consider is whether a substantial risk of error exists. A substantial risk of error will exist if, and only if, the Relevant Sphere contains worlds in which the belief is false. In cases in which there is no substantial risk of error – i.e. when the Relevant Sphere contains no worlds in which your belief is false, the belief will satisfies the ROE constraint by default. No additional condition will have to be met for the constraint to be satisfied.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

Notice that the ROE constraint is even less demanding in these cases than DeRose's notion of an epistemic position strong enough for knowledge in everyday contexts. For, according to DeRose, even if the Relevant Sphere contains only worlds in which your belief is true, your epistemic position will fail to be strong (to the degree under discussion) if in some of those worlds you don't have the belief (or at least if you believe its negation).[20] No such restriction is imposed by the ROE constraint. If you believe that p, and p is true throughout the Relevant Sphere, then your belief will satisfy the ROE constraint even if in some of these worlds you don't believe that p.

I think that intuition is firmly on the side of permissiveness on this point. If BIV-worlds are indeed as distant as I think they are, then I can't see why the existence of nearby worlds in which I believe that I am a brain in a vat, as a result of, say, brainwashing, or too much philosophy – why the existence of these worlds should pose an obstacle to bestowing on my actual belief in not-BIV the status of knowledge.

This corresponds to a general difference between the conception of error with which DeRose operates and the conception that I propose to treat as relevant. On DeRose's picture, the risk that needs to be kept at bay in order for knowledge to be possible is the risk that your belief as to whether or not p might be in error. Clearly this risk is posed not only by worlds in which you believe that p but p is false, but also by worlds in which you don't believe that p (or you believe that not-p) and p is true. What I am proposing is that what is relevant to whether or not your belief that p has the status of knowledge is the risk that you might erroneously believe that p. Consequently, the risk that you might erroneously fail to believe p, or believe not-p, will be irrelevant to the satisfaction of the ROE constraint.

This feature of my approach will be reflected by the fact that the ROE constraint will abide by the following *Principle of Asymmetry*:

> **PA:** If S believes that p and p is true, satisfaction of the ROE constraint by S's belief that p will not depend on what happens in counterfactual situations in which p is true.[21]

[20] "Where not-P (here, *I am a BIV*) is quite remote, one can be in a quite strong epistemic position with respect to P merely by believing that P in all the nearby worlds" (Ibid. 35). The point that DeRose is making is that nothing but believing that P in all the nearby worlds is required in these cases for a strong epistemic position, but he seems to be asserting, by implication, that this is a requirement for a strong epistemic position.

[21] It might clarify matters to think that DeRose's account of an epistemic position sufficiently strong for everyday purposes is related to Nozick's account of knowledge as the ROE constraint is related to the account of

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

So far we have considered cases in which no substantial risk of error is present. I have argued that in these cases no further condition is required for the satisfaction of the ROE constraint. So if worlds in which sceptical hypotheses are true are as remote as we think they are, i.e. outside the Relevant Sphere, our beliefs to the effect that they are false will satisfy the constraint in this way.

Let me now turn to cases in which a substantial risk of error is present – beliefs that are false somewhere in the Relevant Sphere. When the risk of error is substantial, satisfying the ROE constraint will require bringing the risk under control. One way in which this can be achieved is by being protected against the risk, and a belief will be protected from the risk of error precisely when it is sensitive. If you believe that p and things are such that, if p were false you wouldn't believe that p, then the risk of error posed by non-p worlds in the Relevant Sphere will be kept at bay.

Notice that, in accordance with the Principle of Asymmetry, it is not required, in addition, that S believes that p in nearby situations in which p is true. What S believes in those situations will have no consequences for whether S's belief that p is protected from the risk of error. This aspect of my approach answers to the intuition that S's true belief that p can have the status of knowledge when its sensitivity is due to the fact that S's cognitive devices make belief in p dependent on the satisfaction of a condition q that isn't satisfied in any not-p worlds but also goes unsatisfied in many nearby p-worlds. If, in these circumstances, S believes that p because her cognitive devices have detected the satisfaction of q, there will be nearby p-worlds in which S doesn't believe p.

This way of satisfying the ROE constraint corresponds to the way in which, in DeRose's picture, we achieve a strong epistemic position with respect to the propositions that figure in our everyday knowledge claims. The construal of the ROE constraint that I want to put forward will differ from DeRose's approach in offering an alternative method for bringing the risk of error under control, in addition to being protected against it. My proposal is that the risk will also be under control when the subject has identified adequate evidence in its support.

For the purposes of the ROE constraint, adequate evidence for p is a true proposition q that wouldn't be true if p weren't true. If we say that fact A is *sensitive* to fact B just in case A doesn't obtain in the nearest worlds in which B doesn't obtain, q will provide adequate

---

knowledge that would result if we removed Condition 4 from Nozick's analysis. Cf. R. Nozick 1981: 176–78.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

evidence for p when the fact that q is sensitive to the fact that p. If q constitutes adequate evidence for p, S will have identified this evidence when she believes that q and that q is sensitive to p, and these beliefs of hers satisfy the ROE constraint. Clearly, this situation will only effect satisfaction of the ROE constraint by S's belief that p in cases in which this belief doesn't independently satisfy the constraint. When satisfaction of the constraint by S's belief that q and that q is sensitive to p somehow presupposes its satisfaction by S's belief that p, S's belief that p will never come to satisfy the constraint through S's identification of the evidential support provided by q. I shall refer to this form of risk control as *evidential control*.

My proposal is, then, that in cases in which there is a substantial risk of error, the ROE constraint will be satisfied either when the belief is protected from the risk of error by sensitivity, or when the subject has identified adequate evidence for it.

4. Evidence and Sensitivity

One obvious difference between this approach and the account of the strength of one's epistemic position advanced by DeRose is that in DeRose's picture there is no analogue of the evidential method for bringing the risk of error under control. DeRose clearly accepts that one's epistemic position with respect to a proposition can be made strong, to the requisite degree, by identifying evidence for it. Nevertheless he sees no need to mention evidence separately as a source of epistemic strength, because he thinks that these cases are already covered by the stipulation that one's epistemic situation can be made strong by sensitivity.[22] The point that I am attributing to DeRose can be expressed in terms of the following principle linking evidential knowledge and sensitivity:

[22] Here is a representative passage: "[. . .] by checking appropriately independent sources, I could get myself into a position in which I seemingly *would* know that the newspaper isn't mistaken about whether the Bulls won last night. But the checks that would seemingly allow this knowledge would also make it seem that if the paper were mistaken, I would *not* believe it wasn't." (K. DeRose 1995: 25). Nozick made a similar claim in his discussion of strong evidence, although Nozick's version of the thought is rendered less vulnerable by the fact that he restricts it to cases in which you believe h *on the basis of* strong evidence e, where this requires that your belief that h "depends upon (and varies with)" your belief that e. See R. Nozick 1981: 249. And, as we are about to see, Nozick, unlike DeRose, explicitly endorses the controversial consequences of his claim.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

ES: S's identification of adequate evidence for p bestows on her true belief that p the status of knowledge just in case it renders the belief sensitive.

I want to defend my proposal that evidence should be treated separately by raising some problems concerning ES.

The source of the problems is that, as it stands, the principle has obvious counterexamples. One is provided by Nozick's grandmother case: "A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset".[23] While we would want to say that the grandmother knows that her grandson is well as a result of the evidence that she identifies during his visit, her belief is not sensitive. The grandmother's belief that her grandson is well is a counterexample to ES.

The strategy that Nozick puts forward to deal with this difficulty, for which DeRose expresses guarded support, is to relativise the notion of sensitivity to the method employed for forming the belief. The grandmother believes that her grandson is well in the nearest worlds in which he isn't well, but in these worlds her belief is not formed with the same method with which she forms it in the actual world. Her belief that her grandson is well can still have the status of knowledge if she doesn't have it in the nearest worlds in which her grandson is not well *and she arrives at her belief whether or not he is well with the method that she used in actuality to form her belief that he is well*.

Simplifying somewhat Nozick's presentation, we can introduce the following method-relative notion of sensitivity:

S's belief that p is *M-sensitive* just in case, if p were false and S were to arrive at a belief whether or not p with the method she actually used for forming her belief that p, she wouldn't believe that p.

We can now use this notion to formulate a version of the Evidence-Sensitivity Principle that addresses the difficulty:

ES*: S's identification of adequate evidence for p bestows on her true belief that p the status of knowledge just in case it renders the belief M-sensitive.

Now we'll be able to say that the grandmother's belief that her grandson is well obtains the status of knowledge from the evidence

[23] R. Nozick 1981: 179.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

that she gathers during his visit provided that she doesn't believe he is well in the nearest world in which he is unwell and she arrives at her belief whether or not he is well by the same method.

But not all cases of evidential knowledge can be handled so easily. The reason why the model is suitable for this case is that the sensory evidence gathered by the grandmother can be described as resulting from the application of a method (call it casual inspection) capable of producing the belief that her grandson is not well as well as the belief that he is well. This enables us to single out the worlds in which she would have to refrain from believing that her grandson is well in order for her actual belief to be M-sensitive – the nearest worlds in which he is not well and she arrives at her belief whether or not he is well by casual inspection.

In other cases in which q provides adequate evidence for p, belief in p cannot be described as resulting from the application of a method for arriving at a belief whether or not p that could also be used if p were false. Consider cases in which q is the positive result of a test with no false positives but lots of false negatives. The worlds that we would need to look at to determine the M-sensitivity of S's belief that p would be the nearest non-p worlds in which S arrives at a belief whether or not p on the basis of that method. But not-p worlds can be expected to be not-q worlds, and in light of the evidential irrelevance of not-q to whether or not p, it is not clear under what circumstances we should say of a not-q world that in this world S has arrived at her belief whether or not p on the basis of the method that she used in actuality for forming her belief that p. Presumably it would have to be a world in which the evidential irrelevance of not-q for the truth value of p is not outweighed by any proposition that S believes and considers evidentially relevant for the truth value of p – i.e. a world in which S regards all the propositions that she believes, including not-q, as evidentially irrelevant to whether or not p. But then describing S's suspension of judgment on p as resulting specifically from the q-method would seem entirely arbitrary.

The case for describing S's counterfactual suspension of judgment on p as resulting from the same method that she actually employed for forming her belief that p is particularly weak in cases in which the serendipitous nature of the evidence that actually lead to belief in p would make it unlikely that, if the evidence didn't obtain, its failure to obtain would even occur to S. Suppose that Mary's son disappeared years ago. One day she finds on the street a copy of today's newspaper with what she conclusively identifies as her son's signature written on it. It is natural to suppose that

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

this discovery can bestow on her belief that her son is alive the status of knowledge. But accounting for this knowledge in terms of ES* would require considering situations in which Mary refrains from believing that her son is alive as a result of her realisation that she hasn't found on the street a copy of today's newspaper with her son's signature on it. And it is hard to see how this realisation could ever be the main factor in Mary's counterfactual decision not to believe that her son is alive.

The general point that these observations illustrate is that the way in which evidence can confer on S's true belief that p the status of knowledge can't always be naturally characterised in terms of the application of a method for arriving at a belief as to whether or not p. But describing evidence in these terms is unavoidable as soon as we decide to characterise the way in which evidence can confer on a belief the status of knowledge in terms of the sensitivity of the belief.

My proposal is, then, that the risk of error regarding S's true belief that p can be brought under control either by the protection that results from belief sensitivity or by the assurance that error is not present afforded by evidence. In normal circumstances, both forms of control go together. On the one hand, if the facts about the subject's cognitive devices that make her belief sensitive are known to her, they will provide her with adequate evidence of its truth. On the other hand, when S's belief that p is based on adequate evidence, normally it will also be sensitive: if p were false, the evidence wouldn't obtain, the subject wouldn't believe in the evidence, and the subject wouldn't believe that p. Nevertheless, either form of control can in principle be present in the absence of the other. We have just seen how evidence without sensitivity can arise. Sensitive belief in the absence of evidence is also a possibility, so long as our notion of evidence incorporates even a minimal accessibility constraint. Cases of this kind are provided by beliefs whose sensitivity results from reliable sub-personal belief-forming devices of which the subject cannot be aware without sophisticated scientific research.[24] What I am proposing is that either form of risk control will suffice on its own to satisfy the ROE constraint.

[24] Notice that a constraint that requires evidence whenever there is a substantial risk of error will be too strong if sensitive belief is regarded as an adequate form of risk control. My principle EW (see my "Externalism, Skepticism and the Problem of Easy Knowledge") would have to be modified accordingly.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

5. Risk, Knowledge and Scepticism

We can now use the ideas I have presented to provide a formulation of the ROE constraint:[25]

> **ROE:** If S believes that p and p is true, then S knows that p only if either p is true throughout the relevant sphere or S's belief that p is sensitive or S has identified adequate evidence for p.[26]

Notice that the ROE constraint is only a necessary condition for knowledge. It's certainly not universally sufficient. All beliefs in necessary truths satisfy the ROE constraint trivially. A similar point applies to belief in true natural laws.[27] Knowledge in these cases might require evidence or sensitivity even if the ROE constraint doesn't call for it. Nevertheless, I want to put forward the hypothesis that these are the only cases in which the situation might arise. In all other cases, knowledge requires evidence or sensitivity only if the ROE constraint calls for it. I am going to refer to this hypothesis as the *Limitation Clause* (LC).

I want to turn now to considering how ROE (and LC) can be used for dealing with the standard lines of reasoning in support of premise 1 of the canonical argument – the claim that we don't know that sceptical hypotheses don't obtain. The claim has been defended by two different routes, corresponding to the two forms of control of the risk of error contemplated by the ROE constraint.[28] Some have argued, on the one hand, that the reason why I don't know not-BIV is that I don't have adequate evidence in support of this proposition. Furthermore, this predicament can't be overcome, since the BIV hypothesis is precisely designed so that no evidence that I might conceivably gather could support its

[25] The set of beliefs that satisfy the ROE constraint for a person at a time can be inductively defined. The base will contain those true beliefs of the subject which are either sensitive or true throughout the Relevant Sphere. And the inductive clause will stipulate that if the set contains S's belief that q and S's belief that q is sensitive to p, then it also contains S's belief that p.

[26] There are important connections between the ROE constraint and the notion of safety used in some recent accounts of knowledge. See E. Sosa 1999, D. Pritchard 2005. See also Tim Williamson's notion of *safety from error*: T. Williamson 2000: 123–31.

[27] Thanks to Ciara Fairley for helping me see this.

[28] For a discussion of the relative merits of these strategies, see A. Brueckner 1994: 828–30.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

negation.[29] Others have argued, on the other hand, that I don't know not-BIV because my belief in this proposition is insensitive – if it were false, if I were in fact a brain in a vat, I would still believe that I'm not.[30]

I think that the basic premise of each of these arguments has to be conceded to the sceptic. My belief in not-BIV is certainly not sensitive. And I can't obtain adequate evidence in its support.[31] It follows that the risk of error of my belief in not-BIV is not controlled either by sensitivity or by evidence. However, armed with the ROE constraint (and LC), we can object to the transition from each of these premises to the conclusion that we don't know not-BIV. The reason is obvious. From the fact that my belief in not-BIV is insensitive, and its risk of error is not evidentially controlled, it follows that the risk of error is not under control.[32] But lack of control is a problem only when the risk is substantial. And in the case of sceptical hypotheses, the risk is not substantial – the Relevant Sphere contains no worlds in which they are true. This means that my beliefs to the effect that they don't obtain satisfy the ROE constraint even though their risk of error is not under control. They satisfy the ROE constraint by virtue of the sheer remoteness of the worlds in which they are false. As far as the ROE constraint goes, I know not-BIV even though my belief is insensitive and its risk of error is not evidentially controlled. I know this 'by default' – because things would have to be radically different from the way they are in order for my belief to be false.

[29] This is the strategy for supporting premise 1 of the canonical argument endorsed by Dretske. See F. Dretske 1970: 1016.

[30] This is the strategy adopted by Nozick (see R. Nozick 1981: 200–3). Notice that, as Brueckner has pointed out, someone who, unlike Nozick, endorses the canonical argument, would be ill-advised to defend its first premise in this way, as the claim that knowledge requires sensitivity would undermine the Closure principle, which is the main source of support for the second premise of the argument (see A. Brueckner 1994: 828).

[31] This point has been contested. See J. Pryor, "The skeptic and the Dogmatist". Arguing that I cannot obtain evidence for the conclusion that sceptical hypotheses don't obtain would require imposing additional conditions on the possession of adequate evidence. I have discussed this issue in "Wright on Moore".

[32] Notice that neither argument on its own would suffice to establish this. Both lack of evidential control and insensitivity would have to be invoked in order to obtain the conclusion.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

Of course we might be wrong about this. Worlds in which the sceptical hypotheses are true might be much closer that we think they are, and the Relevant Sphere might contain some of them. And, needless to say, the actual world might be such a world. If the sceptic could provide adequate support for these claims, we would have to accept her conclusion. But the sceptic won't expect to have much success through this route. Her hope was to show that, even if things were as we believe them to be, and even if they had to be as different as we think they would have to be in order for sceptical hypotheses to be true, our beliefs would still not have the status of knowledge. The sceptic's argumentative repertoire contains no resources for establishing that sceptical hypotheses are true either in the actual world or in nearby worlds.

In conclusion, according to the ROE constraint (and LC), it follows from the remoteness of worlds in which sceptical hypotheses are true that I need neither evidence nor sensitivity in order to know that they don't obtain. The remoteness of these worlds is an empirical hypothesis which might turn out to be false, but the sceptic has no argument against it. Therefore, it follows from the ROE constraint (and LC) that the sceptic has no cogent argument for the conclusion that I don't know that sceptical hypotheses don't obtain. I believe I do, and the sceptic's arguments give me no reason to abandon my belief.

6. Pryor's Dogmatism

DeRose uses the label *Moorean* for positions that seek to resist the canonical argument by rejecting its first premise.[33] I have argued that there is a Moorean core in DeRose's contextualist response to the canonical argument. Then I have put forward an invariantist version of the Moorean response. There are other positions in the recent literature that also answer to this description. One that enjoys special prominence is Jim Pryor's dogmatism.[34] In this section I'd like to spell out briefly how the view that I am putting forward is related to Pryor's.

There is one fundamental point on which my position agrees with Pryor's. As fellow Mooreans, we both think that we can resist the sceptic's contention that we don't know that sceptical hypotheses don't obtain. The point at which our views come apart is in our

[33] See K. DeRose 1995: 41.

[34] See J. Pryor 2000. For another proposal along these lines, see T. Black 2002.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

explanations of how this knowledge is possible. One way of bringing out the difference is to consider an argument for the conclusion that I don't know not-BIV which plays an important role in Pryor's discussion, even though he doesn't formulate it in exactly these terms:

A. You can't have knowledge of not-BIV that doesn't rest in part on things that you know by perception.[35]
B. You can't have knowledge of not-BIV that rests in part on things that you know by perception.

Therefore:

C. You can't know not-BIV.[36]

Pryor's strategy for resisting the conclusion of this argument consists in rejecting B. He argues that we know certain things by perception prior to knowing not-BIV, and that knowledge of not-BIV can then rest on these things that we know by perception. In this way, Pryor can satisfy the constraint on knowledge of not-BIV imposed by A, which he finds very plausible.[37]

It seems to me that intuition is firmly against the thought that knowledge of not-BIV can rest on things that you know by perception – e.g. that you have hands or that it's raining. Hence, in my view, the fact that a position explains knowledge of not-BIV in these terms, as Pryor's does, should count, other things being equal, as a reason for rejecting the position. In any case, this is not the place to assess Pryor's rejection of B.[38] My goal is to explain how his strategy differs from mine. And the difference is that, on my position, the way to deal with the argument under discussion is to reject premise A. According to this strategy, the contention expressed by premise B – that my knowledge of not-BIV cannot rest on things that I know by perception – doesn't threaten my knowledge of not-BIV. For this knowledge doesn't have to rest on other knowledge. It results, as far as the ROE constraint goes, from the fact that there are no BIV worlds in the Relevant Sphere.[39] Hence the position that I'm advocating enables us to subscribe premise B, and I regard this, as stated above, as a distinct advantage of this position over Pryor's.

[35] A claim of this form (substituting the evil-demon hypothesis) is equivalent to premise (5) in Pryor's article (2000: 524).
[36] Cf. claim (8) in Pryor's article (Ibid. 528).
[37] Cf. Ibid., 529.
[38] I address this question in Zalabardo *forthcoming*.
[39] I think this view is different from each of the positions opposed to A (Pryor's premise (5)) that Pryor considers (cf. J. Pryor 2000: 524).

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

7. A Revised Proposal

I think that the account presented in previous sections provides a good match for our intuitions concerning propositions that are false in fairly close possible worlds (e.g. I have a broken fingernail) and for propositions that are false only in very remote possible worlds (e.g. I am not a brain in a vat). And these are the kinds of proposition that figure in the canonical argument. However the nearest possible world in which a proposition is false can be at any distance from the actual world between these two extremes. And for propositions for which this distance falls towards the middle of these two extremes, the results are less satisfactory.

Let me refer to the distance between the actual world and the nearest world in which p is false as *p*'s *falsehood distance*. According to our formulation of the ROE constraint, the level of risk control required for your belief that p to have the status of knowledge is a function of p's falsehood distance – but the function only yields two values: zero for falsehood distances greater than the radius of the relevant sphere, and maximum for all the rest. This means that significant differences between the falsehood distances of propositions (e.g. a minimal falsehood distance vs. one only marginally smaller than the radius of the relevant sphere) will have no effect on what level of risk control is required for knowledge. And very small differences (e.g. between falsehood distances marginally smaller and marginally greater than the radius of the relevant sphere) will have a huge effect – the difference between needing full control and needing no control at all.

These anomalies are brought to the fore by some cases that are discussed in the literature on the canonical argument. Consider, e.g. the scenario presented by Dretske in his argument against Closure, in which someone, call her Naari, comes to believe that the animals in a zoo enclosure are zebras by looking at them.[40] Intuitively we might want to say that this belief has the status of knowledge. However, the following argument would seem to rule out this claim:

> Naari doesn't know that the animals in the zebra enclosure are not mules cleverly disguised by the zoo authorities in order to look like zebras.
>
> If Naari doesn't know that the animals in the zebra enclosure are not mules cleverly disguised by the zoo authorities in order to look like zebras, then she doesn't know that they are zebras.

[40] See F. Dretske 1970.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

Therefore:

Naari doesn't know that the animals in the zebra enclosure are zebras.

If we wanted to apply to this argument the strategy that I have presented for the canonical argument, we would need to maintain that Naari's belief that the animals are not cleverly disguised mules satisfies the ROE constraint. But this doesn't seem very plausible. Notice, first, that if we said that the falsehood distance of the proposition that the animals are not cleverly disguised mules (∼M) is greater than the radius of the Relevant Sphere, it would follow that Naari's belief in ∼M would satisfy the ROE constraint even if its risk of error were completely uncontrolled. And this doesn't seem right. Intuition dictates, or so I will assume, that knowledge requires some control in this case. From this we seem forced to conclude that the falsehood distance of ∼M is less than the radius of the relevant sphere. But this means that the ROE constraint will require as much risk control in this case as in any other, and the level of risk control present in this case would be clearly insufficient in other cases. Naari's belief is clearly not sensitive, and she wouldn't normally have adequate evidence either. She has some evidence of a vague sort (e.g. zoo authorities don't tend to do that sort of thing), but it clearly doesn't meet the standards of the notion of evidence that I have presented.

I find this outcome counterintuitive. The falsehood distance of ∼M is intermediate between the falsehood distances of ∼BIV, on the one hand, and of the proposition that I have a broken fingernail, or that the animals in the enclosure are zebras, on the other. The difference between how things are and how they would have to be in order for the enclosure to contain convincingly disguised mules is much greater than the difference between how things are and how they would have to be in order for the enclosure not to contain zebras, but much smaller than the difference between the way things are and the way they would have to be in order for me to be a brain in a vat. I think this fact should be reflected in what level of control of the risk of error would be required in order for Naari's belief in ∼M to satisfy the ROE constraint: some control should be required but not as much as with propositions with much smaller falsehood distances. In general, the level of risk control required should vary gradually with the falsehood distance of the proposition in question. In the remainder I want to put forward a revised formulation of the ROE constraint that satisfies this desideratum.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

The revised proposal will make use of propositions to which I shall refer as *probabilistic counterfactuals* – counterfactuals with a probabilistic conclusion, e.g. *if Tony Blair hadn't won the last election, fox hunting would probably still be legal*. These counterfactuals sustain comparisons of probability. We can say, e.g. *if Tony Blair hadn't won the last election, fox hunting would be more likely to be legal than smoking cannabis would be*. Idealising from these comparisons, we can assign for all facts A and B a numerical value between 0 and 1 to the probability that A wouldn't obtain if B didn't obtain, represented as $CProb(\sim A/\sim B)$.[41] We can refer to this value as A's *sensitivity* to B, and say that A is *k-sensitive* to B when $CProb(\sim A/\sim B) = k$. This notion enables us to specify a continuously variable degree of control of the risk of error as a necessary condition for knowledge. The risk of error of S's belief that p will be controlled to a degree k between 0 and 1 just in case either S's belief that p is k-sensitive (to p) or S has identified k-sensitive evidence for p.

To complete the model, we just need to determine what degree of control will be required for a given belief that p to satisfy the ROE constraint. My proposal is that this will be determined by the falsehood distance of p. Thus we postulate the *control-requirement* function, *cr*, pairing each true proposition p with the number between 0 and 1 that represents the level of control of the risk of error that would be required in order for a belief that p to satisfy the ROE constraint. In order to play this role, the control-requirement function will have to satisfy a few basic conditions. First, it will have to assign the same value to propositions with the same falsehood distance. Second, if the falsehood distance of p is greater than the falsehood distance of q, then p won't receive a greater value than q. Finally, we can add a condition that calibrates the function to the Relevant Sphere: a proposition will have to receive a non-zero value if and only if its falsehood distance is smaller than the radius of the Relevant Sphere.

Using the control-requirement function, we can now provide our new formulation of the ROE constraint:

> **ROE*** If S believes that p and p is true, then S knows that p only if either S's belief that p is cr(p)-sensitive (to p) or S has identified cr(p)-sensitive evidence for p.

Let me close by considering briefly how this account would enable us to deal with Naari's belief that the animals in the enclosure are not

[41] Nozick uses these counterfactuals in his account of evidence based on probability. See R. Nozick 1981: 251–63.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

cleverly disguised mules. In light of our previous discussion of the case, we should expect *cr* to assign a low but non-zero value to the proposition that the animals are not cleverly disguised mules. This means that the ROE constraint will require a certain level of risk control in this case, but not as much as in other cases, e.g. her belief that the animals are zebras or my belief that I have a broken fingernail. Hence it is possible in principle that the weak evidence that she has for ∼M provides her belief with a sufficient level of risk control. This would be so if the probability that her evidence didn't obtain if the animals were cleverly disguised mules is higher than the (low) value that *cr* assigns to this proposition. Hence, while yielding the same results as our previous proposal for propositions with very large or very small falsehood distances, the present account appears to have the resources to provide a more satisfactory treatment of the intermediate cases.

University College London

References

Black, Tim. (2002) "A Moorean Response to Brain-In-A-Vat Scepticism." *Australasian Journal of Philosophy* **80**, 148–63.

Brueckner, Anthony. (1994) "The Structure of the Skeptical Argument." *Philosophy and Phenomenological Research* **54**, 827–35.

DeRose, Keith. (1995) "Solving the Skeptical Problem." *Philosophical Review* **104**, 1–52.

———— (1999) "Contextualism: An Explanation and Defense." In *The Blackwell Guide to Epistemology*, (eds.) Ernest Sosa and John Greco, 187–205. Malden, Massachusetts and Oxford: Blackwell.

———— (2004) "Sosa, Safety, Sensitivity, and Skeptical Hypotheses." In *Ernest Sosa and His Critics*, (ed.) John Greco, 22–41. Malden, Massachusetts and Oxford: Blackwell.

Descartes, René. (1984) "Meditations on First Philosophy." In *The Philosophical Writings of Descartes. Volume II*, (eds.) John Cottingham, Robert Stoothoff and Dugald Murdoch, 1–62. Cambridge: Cambridge University Press.

Dretske, Fred. (1970) "Epistemic Operators." *Journal of Philosophy* **67**, 1007–23.

Nozick, Robert. (1981) *Philosophical Explanations*. Cambridge, Massachusetts: Harvard University Press.

Pritchard, Duncan. (2005) *Epistemic Luck*. Oxford: Clarendon.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press

Pryor, James. (2000) "The Skeptic and the Dogmatist." *Noûs* **34**, 517–49.
Putnam, Hilary. (1981) *Reason, Truth and History*. Cambridge: Cambridge University Press.
Sosa, Ernest. (1999) "How to Defeat Opposition to Moore." In *Philosophical Perspectives*, **13**, *Epistemology*, (ed.) James E. Tomberlin, 141–53. Malden, Massachusetts and Oxford: Blackwell.
Williamson, Tim. (2000) *Knowledge and Its Limits*. Oxford: Oxford University Press.
Zalabardo, José L. (2005) "Externalism, Skepticism and the Problem of Easy Knowledge." *Philosophical Review* **114**, 33–61.
———— (*forthcoming*) "Wright on Moore." In *Wittgenstein, Epistemology and Mind. Themes from the Philosophy of Crispin Wright*, (ed.) Annalisa Coliva. Oxford: Oxford University Press.

https://doi.org/10.1017/S1358246109000071 Published online by Cambridge University Press