

On the self-locating response to the knowledge argument

Daniel Stoljar

© Springer Science+Business Media B.V. 2010

1. Boiled down to essentials, the knowledge argument against materialism (KA) is this:

Knowledge Argument (simple version):

K1. It is possible for someone to know all the physical facts and not know some phenomenal facts (e.g. the fact one knows when one knows what it is like to see green.)

KC. Materialism is false

In this argument, the premise is supported by thought experiments with which I will presume familiarity—Jackson’s Mary is the most famous (see Jackson 1982)—and the conclusion seems to follow if materialism is the thesis that every fact is a physical fact.¹ An argument that is analogous to KA is this:

Analogous Argument:

A1. It is possible for someone to know all the physical facts and not know some self-locating facts (e.g. the fact one knows when one knows who or where one is.)

AC. Materialism is false.

In this argument, the premise is supported by two things: (a) thought experiments with which I will also presume familiarity—Perry’s Stanford amnesiac is probably the most famous (see Perry 1977)—and (b) the idea that physical facts are objective

¹ This is the way Stalnaker interprets materialism. I doubt materialism is quite this thesis but this complication does not matter in what follows. (A more complicated version of materialism would require a more complicated version of KA; that is why the version of KA discussed here is a simple one.)

facts. Is the analogous argument (AA) persuasive? Most people think not. If they are right, and if AA really is analogous to KA (and in relevant respects), KA is unpersuasive too.

This is *the self-locating (or indexical) response* to the KA—so-called because the facts that one is said to be ignorant of in AA are self-locating or indexical facts. It has been discussed by a number of philosophers including Perry (positively; see Perry 1999) and Lewis (negatively; see Lewis 1988).² Now Robert Stalnaker has defended it at length in *Our Knowledge of the Internal World*.³ I think Stalnaker has provided by far the most sophisticated account of the response, but I will suggest that it unfortunately remains implausible as a response to KA.

2. KA and AA are analogous to each other in potentially infinite ways. The first thing any proponent of the self-locating response needs to do, therefore, is explain in what way precisely they are analogous. An obvious suggestion here is that AA and KA are analogous because phenomenal knowledge *just is* a sort of self-locating knowledge. This is the suggestion considered by Lewis in his (1999) discussion of the self-locating strategy. He imagines a critic of KA pointing out that Mary apparently learns two things on coming out: first, what it is like to see green; second, that she herself is seeing green. Lewis further imagines the critic arguing that these two facts are one and the same. Oversimplifying, we might capture this by construing Lewis's imagined critic as advancing (1):

- (1) The fact that Mary knows when she knows (on the occasion of her coming out of her room) that *she herself* is seeing green = the fact that Mary knows when she knows what it is like to see green

If (1) is true, phenomenal knowledge just is a sort of self-locating knowledge. In consequence, KA is relevantly analogous to AA and is therefore unsound.

However (1) is not true. (a) It is possible for Mary to remember the LHS fact here and not remember the RHS fact (and vice versa); (b) it is possible for Mary to anticipate the LHS fact and not anticipate the RHS fact; (c) it is possible for Mary to communicate the LHS fact and not the RHS fact. It is for these reasons (especially the first) that Lewis rejected the self-locating response.

Now, Stalnaker does not defend this version of the self-locating response and so is not subject to these objections. What then is his version? Well, if we interpret him as advancing a version of the self-locating strategy analogous to the one considered by Lewis, we may interpret him (again: oversimplifying) as advancing (2):

- (2) The fact that Mary knows when she knows (as she would put it on the occasion of her coming out of her room) that *this* is what it is like to see green = the fact that Mary knows when she knows what it is like to see green

Once again, if (2) is true, KA is unsound. Before assessing this proposal, let me bring out two features that distinguish it from the proposal considered by Lewis (apart of course from the self-locating facts involved).

² See also Chalmers 2004.

³ Stalnaker's goals are considerably more ambitious than simply providing a response to KA. But I will focus here on what he says about KA.

3. Stalnaker suggests the LHS fact in (2) (and so the RHS fact too) is an *essentially contextual fact*. To bring out the idea, Stalnaker give us the example Barry and Alice:

Alice, of the Homeland Security bomb squad, is in the Rose Garden on Tuesday morning, 10:47 am. She points to the ground beneath a particular rose bush, and says, “a bomb is buried there, and unless we defuse it now it will explode in five minutes.”

Barry is in a room far away from the Rose Garden on Monday, and he knows that the next day, at 10:47 am, there will be a bomb buried under the rose bush which is 10.25 meters east and 4.35 meters north of the southwest corner of the garden (the very place that Alice will be pointing to), and that unless it is defused soon after that, it will explode before 10.52. (2008, p.85)

“Barry knows a lot about the situation”, Stalnaker goes on, “but he doesn’t know what Alice will know the next day. He is not in a position to know this; he would have to be there” (p. 85). The proposition that Alice expresses is essentially contextual, in that it “cannot be detached from the context in which it is expressed” (p. 85).

The Barry/Alice example is interesting in a number of ways. For one thing, it illustrates something that for Stalnaker is the main lesson of the Perry-style examples, viz., the existence of essentially contextual facts. For our purposes, however, the main point of the example is that Barry’s position is analogous to pre-release, and Alice’s to post-release, Mary; hence the facts mentioned in (2) are essentially contextual. This constitutes the first major difference between his proposal and the one discussed by Lewis, for whom the notion of an essentially contextual fact plays no role.

4. The second difference is that for Stalnaker what Mary learns when she comes out is the truth of a possible worlds proposition, i.e., a proposition that distinguishes one set of possible worlds from another. This marks a difference between Stalnaker’s proposal and the one discussed by Lewis. For it is crucial to the proposal discussed by Lewis that what Mary learns is the truth of a centred worlds proposition, i.e., a proposition that distinguishes a set of centred worlds (i.e., pairs of ordinary possible worlds and individuals) from another.

Why this difference is important? Well, that Mary learns a (genuinely) centred worlds proposition is crucial to the idea discussed by Lewis being plausible in the first place. If there are both (genuinely) centred worlds propositions and possible worlds propositions in play, it is fairly clear (though I will pass over the details here) that the AA is fallacious, for the quantifiers contained in the argument (‘all the physical facts’, ‘some self-locating fact’) may be interpreted as sometimes restricted to one sort of proposition and sometimes not. But Stalnaker disowns this distinction (at least in this context). Why then *is* AA (and indeed KA itself) fallacious according to Stalnaker?

It is important to note that appealing to the idea that the facts in (2) are essentially contextual does not answer this question. After all, one can easily enough imagine a dualist saying: “OK perhaps the proposition that Mary learns *is* an essentially contextual proposition—but so what? So long it is a possible worlds proposition

and so long as she already knew all the physical possible worlds propositions beforehand, this is sufficient for my purposes.”

So far as I can see (though this is a speculation)⁴ what Stalnaker needs to argue here is that Mary does *not* know all the physical facts prior to coming out, and that we are somehow mistaken when we describe her in that way. What she perhaps knows are all the physical *facts that it is possible for her to know while in her room*, but not all the physical facts strictly and literally. To interpret Stalnaker this way is to view him as holding what I have called elsewhere the ‘epistemic response’ to the argument. This is, as I say, a speculation: it is unclear to me whether Stalnaker intends this commitment. But it does seem to be a natural interpretation of what he says.

5. So much for setting out Stalnaker’s view—is it successful? Well if the fact mentioned in the LHS of (2) is essentially contextual, the view does seem an advance of the proposal discussed by Lewis at least as far as the points about anticipation and communication goes. But—and this is the first of two objections I want to raise—the proposal still faces a problem about memory.

Suppose that Mary comes out of her room at 9 am sharp on Monday morning. At that point she learns (as she would put it) that *this* is what it is like to see green and (if this is different) what it is like to see green. Now, Stalnaker quite naturally wants to allow for the possibility that later (on Tuesday, for example) Mary could remember what she learned on Monday. However—and here is the objection—offhand it seems perfectly possible that Mary could on Tuesday *remember* that this is what it is like to see a green thing, and nevertheless *forget* what it is like to see a green thing. Suppose that sometime during Monday evening, Mary suffers from partial amnesia, and forgets what is like to see a green thing. It is certainly not obvious that she would *likewise* forget (as she would continue to put it) that that was what it is like to see green. Imagine that on Tuesday morning, she is trying hard to remember what seeing green is like, perhaps with accompanying feeling that if she only concentrated, she could succeed in recalling it. She might say, “Gosh, I remember that that was what seeing green is like, but I can’t remember what seeing green is like.”

One might respond that this is not possible. But at least *prima facie* this is not so, and there are examples from ordinary life that support it. Suppose I am at the paint shop discussing with the assistant which colour I should paint my feature wall. “This is the colour you should use,” he says pointing out some particular shade, and I find myself agreeing. On my way home I suffer from partial amnesia. I retain my ability to demonstratively refer to the shade that the assistant was trying to convince me to adopt but lose my ability to remember what that shade is like. I might in that context say, “I remember that *that* was what the colour is like, but I can’t remember what the colour is like!” This is certainly a frustrating state of affairs but it does not

⁴ I am indebted to Paolo Santorio for this suggestion. An alternative suggestion might be to construe physicalism as the thesis that every fact is either physical or is entailed by some physical fact, and then say that while Mary knows all the physical facts she does not know all the facts entailed by a physical fact. I think this will ultimately resolve itself into a version of what Stalnaker calls (and rejects under that name) the Fregean Strategy. In any case I will set it aside here.

seem impossible. But, if I understand him correctly, it should be impossible if Stalnaker's proposal is right.⁵

Alternatively, one might respond that, while this objection shows that identity statements such as (2) are false, it does not show that Stalnaker's proposal is mistaken. So far I have been assuming that Stalnaker is committed to (2) because I have been assuming that for him, like Lewis's imagined critic, phenomenal knowledge just is self-locating knowledge. But perhaps this is mistaken. Instead of defending the analogy between AA and KA by saying that phenomenal knowledge just is self-locating knowledge, Stalnaker might instead defend it by saying that both phenomenal and self-locating knowledge are essentially contextual. However, while this suggestion might avoid my first objection—the one about memory—it does not avoid the second, which I will now go on to consider. According to this second objection, phenomenal knowledge is not essentially contextual, or at least is not in the way the Stalnaker intends.

6. When Mary comes out of her room and says “this is what it is like to see green”, the property she demonstrates is a physical property of a person, a property that a person instantiates when they see green; that much follows from physicalism of the sort Stalnaker has in mind. Now, this property is certainly one that Mary could have singled out in thought, demonstrated, and talked about prior to coming out. She might have said while in her room “that property is what causes people to say the English sentence ‘that’s green’”. Moreover, she could certainly have raised the possibility that that property is what it is like to see green. Imagine a dialogue between Mary and one of her co-workers about whether that is what it is like to see green. The co-worker says “Here is a speculation: *that* is what it is like see green.” Mary (the more cautious one) replies as she steps through the door, “We will see in a moment”. When she comes out a moment later, she immediately calls back to the co-worker: “You were right: *that is* what it is like to see green!” The fact Mary expresses knowledge of when she says this is very plausibly the very same fact that the co-worker speculated to be the case a moment previously. If so it is quite unclear that Mary's knowledge is essentially contextual in the way that Stalnaker says.

The problem for Stalnaker's proposal here seems to me a serious one because it reveals a significant relevant dis-analogy between the Mary case and the Barry/Alice case. The fact that Alice learnt could not have been entertained by Barry, but the fact that post-release Mary learnt could have been entertained by pre-release Mary (e.g., she can wonder whether it obtains); it is simply that she doesn't know whether that fact obtains. More generally, when Stalnaker says that pre-release Mary, like Barry, is not in a position to know some relevant facts, there are two things he could mean. First, he could mean that she was not in a position to know the relevant facts because she could not so much as entertain them (i.e., she could not wonder whether they obtain, speculate that they obtain and so on). Second, he could mean that she was not in a position to know the relevant facts because, while she could entertain them, her epistemic situation did not allow her to know them. As

⁵ A different example, though a more distant one, might be this. At ANU we get a lot of visitors, and, as head of department, I have to sign lots of forms with the names of various philosophers on them. Sometimes I can remember seeing the name of a particular philosopher but can't remember what the name of the philosopher is. I might say, “I remember that that was the name, but I can't remember what the name is.”

I understand it, appealing to the notion of a proposition's being essentially contextual involves the first idea. But the KA crucially relies on the second. But if that is so, then the facts that Mary learns are not essentially contextual in the way that Stalnaker intends.⁶

7. I have suggested that two versions of the self-locating response to the KA are unsuccessful: the one considered by Lewis, the other advanced by Stalnaker. Could anything here be successful? I will close by looking briefly at an argument that Stalnaker gives (as I read him)⁷ which suggests that the self-locating response *must* be right in outline, even if we have not found its right form.

The argument begins with two observations. The Lewis observation (as I will call it) is this:

Let parapsychology be the science of all the non-physical things, properties, causal processes, laws of nature, and so forth that may be required to explain the things we do. Let us suppose that we learn ever so much parapsychology. It will make no difference. Black-and-white Mary may study all the parapsychology as well as all the psychophysics of colour vision, but she still won't know what it is like...If there is such a thing as phenomenal information, it is isn't just independent of physical information. It's independent of every sort of information that could be served up in lessons for the inexperienced...Therefore phenomenal information is not just parapsychological information, if such there be. It is something much stranger. (Lewis 1988, p. 280 quoted in Stalnaker 2003, p. 270)

The Nagel observation (as I will call it) is this:

If we suppose 'being me' to be any objective property whatever of the person TN, or any relation of that person to something else, the supposition quickly collapses. We are bound to include that property or that relation in the objective conception of the world that contains TN. But as soon as it has been made an aspect of the objective TN, I can ask again, 'Which of these persons am I?' and the answer tells me something further. No further fact expressible without the first person will do the trick. However complete we make the centerless conception of the world, the fact that I am TN will be omitted. (Nagel 1986, p.56 quoted in Stalnaker 2003, p. 260)

As I read him, Stalnaker suggests that there is a unifying theme here. The theme is (roughly) this: expanding one's ontology to deal with puzzles about phenomenal or self-locating knowledge is pointless (or anyway is pointless without adopting an

⁶ One disanalogy is that the Alice/Barry example is about entertainment, while the Mary example is not. Another is that what is at issue in the Mary example is a property; that is, the expression 'this' in (2) denotes a property. The Alice/Barry example, by contrast, concerns a particular, i.e., a place, rather than a property. But it is not quite clear to me how to extend the notion of an essentially contextual proposition from a particular to a property. This point is also made in Weber 2008.

⁷ Admittedly this idea seems to me to be more clearly on show in Stalnaker 2003 than Stalnaker 2008.

independently objectionable epistemology—I will set this complication aside here.) Since it is pointless to expand one’s ontology, the problem must be a problem, not about what the facts are, but about our *representation* of what the facts are; the self-locating response is simply the best way of spelling that out.

This is certainly a seductive line of thought. Both observations are plausible, and it does seem right to say that the Nagel observation shows that expanding one’s ontology is pointless, at least in the sense that it is pointless in response to puzzles about self-location to believe in the *existence* of properties like being me. However, I don’t think the Lewis observation is about the *existence* of phenomenal facts so much as their *fundamentality*, i.e., whether phenomenal facts are metaphysically distinct from every other sort of fact. What the Lewis observation shows, as I understand it, is that it is quite pointless in response to the KA to postulate further non-physical fundamental features of the world (auras, for example) while retaining the idea that ordinary facts like what it is like to see green are non-fundamental. The reason it is pointless is that knowledge of what it is like to see green seems as remote from knowledge of auras as it is from knowledge of atoms. On the other hand, it is not pointless (though it ultimately might be mistaken) to respond to KA by saying that what it is like to see green is a fundamental fact. For suppose ordinary facts like this were fundamental facts but were associated by contingent laws with other fundamental facts about atoms (or auras for that matter). Then it would be quite possible for Mary to work out what phenomenal facts obtain from a description of the physical facts plus the laws. This seems to me quite different from the self-location case in which it is pointless to say that there exists a property of being me of any sort, fundamental or not.

Of course, that it is not pointless to say that phenomenal facts are fundamental does not mean it is right. My own view is that a version of the epistemic response is correct as an account of KA (see Stoljar 2006). Obviously I cannot go into that here. The important point is that, however the epistemic response is developed, it will not rely on an analogy between KA and AA.

Acknowledgments Thanks to an audience at ANU, and to John Broome, David Chalmers, Frank Jackson, Paolo Santorio and Clas Weber for discussion.

References

- Chalmers, D. (2004). Imagination, indexicality and intensions. *Philosophy and Phenomenological Research*, 68(1), 182–190.
- Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.
- Lewis, D. (1988). What experiences teach. In *Proceedings of the russellian society* (pp. 29–57). Sydney: University of Sydney. (Reprinted in Lewis, D. (1999). *Papers in metaphysics and epistemology* (pp. 262–290). Cambridge: Cambridge University Press, References to the latter).
- Nagel, T. (1986). *The view from nowhere*. Oxford: Oxford University Press.
- Perry, J. (1977). Frege on demonstratives. *Philosophical Review*, 86, 474–497.
- Perry, J. (1999). *Knowledge, possibility and consciousness*. Cambridge, MA: MIT Press.
- Stalnaker, R. (2003) On Thomas Nagel’s objective self. In Stalnaker R. (Ed.), *Ways a world might be: metaphysical and anti-metaphysical essays* (pp. 253–275). Oxford: Oxford University Press.
- Stalnaker, R. (2008). *Our knowledge of the internal world*. Oxford: Clarendon Press.
- Stoljar, D. (2006). *Ignorance and imagination*. New York: Oxford University Press.
- Weber, C. (2008). Review of Stalnaker (2008). *Disputatio* Vol. III, No. 25, November 2008, (pp. 59–65).