

## COMMENTS ON SMITHIES

Thomas Raleigh, University of Luxembourg

Declan Smithies' brilliant book *The Epistemic Role of Consciousness* sets a new benchmark for developing a broadly internalist vision of how consciousness, mental content and justification hang together. The book provides a fully worked out philosophical system that encompasses the major traditional questions in both philosophy of mind and epistemology as well as many recent 'hot topics'. I should confess that I had always previously kind of lazily assumed that all the sophisticated cool kids these days were externalists of some kind or other and that internalism was somehow a bit passé. But Smithies shows how it is possible to develop an extremely powerful and coherent internalist picture, based on the very plausible core intuition that conscious experience (both of our surroundings and of our own minds) plays a vitally important and distinctive epistemic role in our mental lives. It has genuinely shifted my own opinions and it will surely shape the debate on these topics for a long time to come.

Like any philosophical system, Smithies' account relies on various assumptions which play a more or less foundational role and which different readers will perhaps find more or less attractive. For example, Smithies endorses the thesis he calls 'Representationalism: All consciousness is a kind of mental representation' (p34). He also endorses 'Uniqueness: Necessarily, there is exactly one doxastic attitude that you have justification to hold toward any given proposition at any given time.' (p226) and denies that there can be genuine rational/epistemic dilemmas. For what it is worth, I have myself argued against representational accounts of perceptual consciousness (Raleigh 2015), against Uniqueness (Raleigh 2017) and proposed a putative rational dilemma (Raleigh 2021). But I suspect it would be boring and unproductive to focus on these disagreements. Instead, my comments on the book are grouped around the following themes:

- 1) ZOMBIES
- 2) PROPOSITIONAL vs. DOXASTIC JUSTIFICATION
- 3) MOOREAN PROPOSITIONS

In his review of the book, Robert Howell commented that 'Frequently, as soon as one thinks of a shortcoming or a flaw in an argument, Smithies sees it himself and replies.' (Howell, 2020) That was certainly also my experience when reading the book and I suspect that Declan will have ready responses to all the issues and questions I raise below. I am greatly looking forward to learning what those responses are...

### (1) ZOMBIES

(i) SCIENTIFIC INDISPENSABILITY: Smithies allows that zombies would have representational mental states. His motivation here is that attributing representational content to unconscious mental states is part of our best scientific theories:

"We should believe there is unconscious mental representation because it plays an indispensable role in psychological explanation both in common sense and in cognitive science." (p48)

This strikes me as entirely sensible. However, I guess I feel that something like the same motivation should also equally support ascribing properties like justification and rationality to zombies. After all, if we are taking the thought-experiment of zombies seriously then we can imagine a parallel world that is physically and causally identical to our own, populated with perfect physical-functional duplicates of us all, who go around speaking exactly like us, behaving exactly like us, etc. but

without any phenomenal consciousness. It seems extremely plausible that in order to understand and predict and make sense of all this zombie behaviour and zombie chatter and inter-zombie interaction etc. our best theories would indispensably appeal to notions of rationality and justification. Rational choice theory in economics and game theory would be just as useful with the zombies as with us. In order to make sense of their speech and behavior we would just have to start classifying some their actions and utterances as justified/unjustified, rational/irrational. (Think of all the zombie lawyers making more or less plausible arguments and all the zombie jurors trying to decide innocence or guilt; think of all the zombie scientists, some of whom are making better use of their evidence than others, etc. And then of course think of all the zombies uttering sentences that at least *sound* exactly like they are talking about justification and rationality!) And after all, if we adopt something like the stance of a radical translator or interpreter trying to make sense of zombie behavior, Davidson long ago argued that we must make the assumption that the subject has largely *rational* beliefs and desires about their environment.

Of course, one could go the route of insisting that these zombies would only be exhibiting justification\* or rationality\*, rather than genuine justification or rationality, given their lack of consciousness. But not only would such a move strike me as pretty desperate and implausible, it seems that by parity one should then also insist that our unconscious states and zombie mental states are merely representational\*. The point is simply that the appeal to scientific indispensability is just as strong for the case of zombies having rational/justified mental states which are part of the best explanation of their behavior as the case for zombies having representational mental states which are part of the best explanation of their behavior.

(ii) BASING PHENOMENAL BELIEFS vs. Z-LIEFS: Secondly, here is a line of thought that comes from the sorts of worries that Shoemaker (1975, 1999) and Kirk (2005) have pressed about zombies.

Though of course there is surely *more* to the basing relation than just causation, it can seem very plausible that it must *at least somehow involve* causation. So for a belief B, to be based on a reason, R, R must somehow or other figure in the causal story of how B was formed or is sustained. Or to put it more generally, for R to be the motivating reason for me to phi, then R must somehow or other figure in the causal story of how I phi. And so in particular when I form a properly based, doxastically justified belief about my own phenomenal properties, these phenomenal properties (or perhaps my conscious awareness of them) are presumably supposed to be (part of) my motivating reason for forming the belief. But now when we consider my zombie physical duplicate, who is in exactly the same physical environment as me, this zombie must have formed its false phenomenal belief/alief/zlief in exactly the same causal way that I formed my true phenomenal belief. So it seems that my phenomenal properties (or my conscious awareness of them) would not after all be playing any causal-explanatory role in the formation of my phenomenal belief. So as long as we are treating zombies as possible, it seems we will be treating phenomenal properties as epiphenomenal and so it is then hard to understand how we can form doxastically justified beliefs about these phenomenal properties given that basing relation requires that they play some kind of causal-explanatory role in the belief formation. (Or to put the same point differently: it is hard to understand how our belief forming processes could be rationally sensitive to the presence or absence of phenomenal properties if these properties make no causal difference to the belief forming processes.)

One possible response to this kind of worry is to deny the assumption that the basing relation must always be causal. This is the route that Chalmers (2003) takes – who insists that the basing relation can instead involve some kind of non-causal acquaintance, and/or that the phenomenal belief is (partially) constituted by the phenomenal properties rather than caused by them. Perhaps Smithies is also happy to go this route? But it strikes me as being at least some kind of theoretical cost insofar as it relies on a somewhat obscure re-jigging of the basing relation.

## (2) PROPOSITIONAL vs. DOXASTIC JUSTIFICATION

(i) FIDO & FIFI: The first point here concerns Smithies' discussion of phenomenal/internal duplicates who are looking respectively at Fido and Fifi, two visually indistinguishable apples<sup>1</sup>. Smithies wants to hold that both duplicates have the exact same propositional justification to form beliefs – so they both have propositional justification to believe (the *de re* proposition) that Fido is rotten AND (the *de re* proposition) that Fifi is rotten. However they cannot both form doxastically justified beliefs in both of these propositions, simply because they cannot both form beliefs in both propositions. For they are perceptually related to only one of Fido vs. Fifi, so they can only doxastically *de re* represent one of these two items.

But so now what about this case? Two phenomenal, internal duplicates TWIN1 and TWIN2 are both looking at two visually indistinguishable scenes. In the scene TWIN1 is looking at, Fido is on the left and Fifi on the right. In the scene TWIN2 is looking at, Fifi is on the left and Fido on the right. Their respective experiences are phenomenally identical. When TWIN1, attending to her experience, forms a demonstrative judgement '*That* is on the left and *This* is on the right' TWIN1 thereby forms a *de re* belief in the proposition that Fido is on the left and Fifi on the right. Whereas TWIN2 forms a *de re* belief that Fifi is on the left and Fido on the right. As they are phenomenal duplicates, Smithies holds that they both have propositional justification to form both of these *de re* beliefs. But notice here that both twins are perfectly capable of forming both these *de re* beliefs as they are both perceptually related to both Fifi and Fido. (As well as attending to the item on the left and thinking "that is on the left" and likewise for the item on the right, they are also perfectly capable of attending to the item on the left and thinking 'That is on the right' and vice-versa.) So it seems that both twins are here able to form both of the conflicting beliefs that they both supposedly have propositional justification to believe. This seems like the wrong result and is conflict with what Smithies labels the Modified Linking principle:

'Necessarily, if you are fully rational, you have sufficient propositional justification to believe that *p*, and you adopt some doxastic attitude toward the proposition that *p*, then you have a doxastically justified belief that *p*.' (p110)

Here, according to Smithies' position, both phenomenal duplicate twins have propositional Justification to believe both the (*de re*) proposition 'Fifi is on the left, Fido is on the right' and the (*de re*) proposition 'Fido is on the left and Fifi is on the right'. And there seems to be nothing that prevents either of them from forming either belief (given they are both perceptually related to both Fido and Fifi). But of course if either twin did form both beliefs, one of those two beliefs would not be doxastically justified.

(ii) RATIONALITY & ATTENTION: Smithies holds that we are always in a position to know exactly what phenomenally conscious states we are in. If we fail to convert this propositional justification into an actual doxastically justified belief it must be due to some departure from the ideal of rationality.

"In some cases, we have introspective reasons that justify believing that we're in some mental state, but we don't succeed in forming a justified belief on the basis of those reasons. An ideally rational agent always converts her epistemic position into knowledge when she forms an opinion on some question. Sadly, we humans are not ideally rational agents...

...[There] are cases in which the subject cannot convert propositional justification into doxastic justification because of some fact about her psychological limitations: she is conceptually impoverished, or inattentive, or indiscriminating, or delusional. As a result, the subject is unable to convert her epistemic position into knowledge." (158)

According to Smithies, an ideally rational agent would not suffer from these limitations. So if she forms a belief at all about her current phenomenal states, she can always pay sufficiently close attention that she gains knowledge.

---

<sup>1</sup> See pages 104-111. The example comes originally from Martin Davies (1997).

OK, so this made me think about cases where a conscious subject has an extremely rich and complex and rapidly changing experience. E.g. suppose that the phenomenal visual field is comprised of an incredibly richly detailed mosaic of coloured dots all of which are moving around very rapidly in incredibly complex motions and also changing colour extremely rapidly. The idea here is that the conscious subject is consciously acquainted with this whole dizzyingly complex shifting mosaic – all of it is genuinely ‘present’ within the stream of consciousness – but when the subject tries to attend to any specific dot or moving feature they find that it is just too swiftly moving and changing to be able to form any demonstrative judgement like – THAT is red, THIS is moving left etc. They cannot attentionally track any dot long or stably enough to be able to form a justified demonstrative judgement about it<sup>2</sup>.

It seem pretty plausible that there are limits to what we normal humans can attentionally track about our own conscious experiences<sup>3</sup>. But so my thought here is simply that this kind of natural limitation on the faculty/ability to selectively attend to and track elements in our stream of consciousness simply does not seem to be any kind of failure of rationality. Rather it seems to be much more closely parallel to natural limitations on our perceptual faculties. As I grow more short-sighted and become less able to visually discriminate fine details or to see very rapidly moving stimuli, this need not amount to any waning of my rationality. And nor is it part of being an ideally rational agent that they would have super-human perceptual powers. But so then why should it cast any shade on my rationality if my faculty of attention likewise has limitations to what it can track and process? Why must an ideally rational agent have super-human powers to selectively attend inwardly anymore than they should have superhuman abilities to perceive outwardly?

I guess I am pressing the intuition here that insofar as we have to selectively attend to features of our own conscious experiences as part of forming justified beliefs about them, it seems that we are not simply gaining perfect, unmediated cognitive/epistemic access to all the phenomenal features in one fell swoop, simply in virtue of having the experience. Rather our cognitive access to our own conscious streams goes via the faculty of attention. And this seems to be at least somewhat like how our cognitive access to the external world goes via perception and via selectively looking at one thing rather than another. And if this faculty of inner attention sometimes provides us with less than full and perfect access to a conscious experience, despite our best efforts to attend as closely and fully as possible, I want to suggest that this is no kind of rational failing. It is more akin to a perceptual limitation.

### (3) MOOREAN PROPOSITIONS

(i) ON MOOREAN FINKISH-NESS: At a number of points Smithies emphasizes that the issue with Moorean propositions such as [MOORE]:  $p$  & I don't believe that  $p$ , is that they are ‘finkish’. In other words he takes the issue here to be that even though it is perfectly possible for your evidence to indicate that the proposition is true, the very act of forming a belief in such a proposition changes your evidence and so the belief ends up being false (and unjustified). Smithies thus insists that you can be ‘in a position to know’ such a proposition, despite the fact that it is necessarily impossible to know such a proposition. This strikes me as a somewhat strange way of putting things (see below). But in any case, what is important is that Smithies thinks one can have propositional justification to believe the Moorean proposition, it is just that one can never convert this to Knowledge due to the Fink-ish-ness of the situation – for the very act of forming the belief changes the evidence and so

---

<sup>2</sup> Some people, e.g. Dennett (1991), would want to deny that there can be phenomenal facts that outstrip even our best attempts to attend to and judge about them. But I take it that this sort of anti-realism about phenomenal facts is not Smithies' position.

<sup>3</sup> There is a whole genre of psychological experiments which study precisely this ability to simultaneously keep track of multiple moving targets. Perhaps the *locus classicus* in this literature is Pylyshyn & Storm (1988).

removes the propositional justification. (And indeed in the case of MOORE it also changes the truth-value of the proposition, assuming it was true to begin with.)

I certainly don't disagree that the situation is Finkish insofar as the very act of forming a belief in MOORE changes your evidence (and indeed can change the truth value of MOORE). However I disagree that *before* a belief in MOORE is formed one can have propositional justification to believe it. For even if your total evidence indicates that MOORE is true, this same total evidence also indicates (in fact entails) that *if* you form a belief in MOORE, then that belief is bound to be false. The fact that you already know (or are in a position to know) this conditional claim is *already* a defeater for your propositional justification to believe MOORE. If your evidence already indicates that B(MOORE) is bound to be false, then you already have a decisive reason not to B(MOORE). The point here is that your evidence not only indicates what is true or false, it also indicates what would be true or would be false if you were to form various beliefs. Both of these kinds of indication bear on what is propositionally justified for you to believe. So I don't think that Finkish-ness is the real issue here. And nor do I think that it is a helpful way of framing things to say that one can be "in a position to know" a proposition like MOORE, just because one's evidence indicates that it is currently true. This does not mean you are in a position to know it, since your evidence also entails that MOORE must be false if believed. So your evidence already contains a defeater *at the level of propositional justification* against believing MOORE.

(ii) IDEAL vs. NON-IDEAL: When discussing Higher-Order (H.O.) evidence and Moorean propositions Smithies distinguishes between what a non-ideal agent (NRA) and what an ideally rational agent (IRA) should do. He holds that a NRA, when faced with misleading H.O. evidence that they are irrational or unreliable etc, cannot reason from the correctness of their 1<sup>st</sup> order belief that p to the conclusion that they can dismiss or ignore the H.O. evidence (a form of reasoning which Smithies calls the 'certainty argument'.) And so a NRA should lower their confidence in p (despite having perfectly good 1<sup>st</sup>-order evidence for p and perhaps previously having a justified belief that p or knowledge that p). Whereas, according to Smithies, an IRA *can* use the 'certainty argument' to dismiss the H.O. evidence and remain fully confident that p. However, Smithies allows that an IRA cannot use the certainty argument to prove that she is an IRA – after all, IRAs are not immune to reason altering drugs or Hypoxia etc. So an IRA's status as an IRA is not luminous in the way that evidential relations and phenomenal mental states are (according to Smithies). Smithies explicitly allows that an IRA can become a NRA (via drugs etc). And so presumably the switch can also occur the other way around, from NRA to IRA (the reason-scrambling drugs wear off, or perhaps reason-enhancing drugs kick in). This all leads Smithies to conclude that for an IRA it can in fact be rational/justified to believe the following Moorean claims:

p and my belief that p is not justified

p and I don't know that p

This is because whilst propositional justification is luminous (according to Smithies), the *basing relation* is not – and neither is one's own status as being an IRA. So an IRA who has very strong evidence that they are *not* an IRA can still tell that p is true (and is supported by the evidence) but they cannot tell that the resulting belief is doxastically justified.

OK, so here is my worry about all this. Suppose that before time t a subject is a NRA and then at t turns into an IRA but without being told or being aware of the change (e.g. they are zapped by a subjectively undetectable rationality ray). Smithies accepts that this difference before and after t is not luminous. Suppose that prior to t the subject always answered questions on some specific topic feeling totally confident that they have correct answer. But they also have strong (misleading) evidence that they are irrational, unreliable reasoners who are always improperly basing their own beliefs. Before t, suppose the subject is asked a question on this topic: is it the case that p? And in fact, after reasoning for a bit, they get the right answer, p is true! But given the evidence about their own irrationality, they should – according to Smithies – lower their confidence in the answer. (They cannot use the certainty argument to dismiss this H.O. evidence since they are merely a NRA). So then after t, once they have been secretly turned into an IRA, they are again asked the question: whether p? At this point according to Smithies, the subject who is now an IRA should believe: p and

I'm not justified in believing that p. For they are certain that p is true and is supported by the evidence. But they cannot be sure they are an IRA and so, given the H.O. evidence that they are irrational, they have to accept that this belief that p is not justified.

But notice: before and after t, when they change from NRA to IRA, *there is no phenomenal, introspect-able difference*. Both times they feel totally confident that p is true and supported by the evidence (and in fact it is!). Both times they have the same evidence that they are irrational, unreliable reasoners. And yet somehow what they are rationally required to do before and after t differ! We could dramatize this by saying that the subject switches constantly every few minutes from NRA to IRA and back again – which would apparently mean they are rationally required to switch back and forth between suspending judgement whether p, to  $B[p \ \& \ \neg JBp]$ . But there is no phenomenally detectable way for the subject to tell when they should make this switch! Apart from just seeming pretty weird, this would also seem to be in tension with Smithies' position that rational requirements are luminous.

### References

- Chalmers, D. (2003) "The content and epistemology of phenomenal belief", in *Consciousness: New Philosophical Perspectives*, edited by Quentin Smith and Alexander Jokic, Oxford: OUP.
- Davies, M. (1997) "Externalism and Experience" in *The Nature of Consciousness: Philosophical Debates*, edited by Ned Block, Owen Flanagan, and Güven Güzeldere, MIT Press.
- Dennett, D. (1991) *Consciousness Explained*, Boston, Toronto, London: Little, Brown
- Howell, R. (2020) Review of *The Epistemic Role of Consciousness* by Declan Smithies, *Notre Dame Philosophical Reviews* [<https://ndpr.nd.edu/reviews/the-epistemic-role-of-consciousness/>]
- Kirk, R. (2005) *Zombies and Consciousness*, Oxford: Clarendon Press.
- Pylyshyn, Z. & Storm, R. (1988) "Tracking multiple independent targets: evidence for a parallel tracking mechanism" *Spatial Vision* 3(3): 1-19.
- Raleigh, T. (2015) "Phenomenology without Representation", *European Journal of Philosophy*, 23(4): 1209-1237
- Raleigh, T. (2017) "Another Argument Against Uniqueness", *The Philosophical Quarterly*, 67(267): 327-346
- Raleigh, T. (2021) "A New Anti-Expertise Dilemma", *Synthese* 199: 5551–5569
- Shoemaker, S. (1975) 'Functionalism and Qualia', *Philosophical Studies*, 27: 291–315
- Shoemaker, S. (1999) 'On David Chalmers's *The Conscious Mind*,' *Philosophy and Phenomenological Research*, 59: 439–444
- Smithies, D. (2019) *The Epistemic Role of Consciousness*, Oxford: OUP.