Comment on Kidd's "The Idols of Inner-Awareness: Towards Disjunctive Self-Representationalism"

Introduction

I welcome the opportunity to comment on Kidd's hypothesis regarding the nature of consciousness and its epistemic implications. In the first section I will raise a concern with his self-representational theory of experience (SR) to the effect that there is, as of yet, no clear way to distinguish the SR account experimentally from its rival HO (higher order) account. I suggest that this is not merely due to a lack of imagination on my part, but rather because the main difference between the two views, namely what counts as good justification for self-knowledge, is inaccessible to the subject and appears to have no possible effect on the actual behaviour of a normal subject. It is a purely normative difference. The challenge to Kidd, then, is to further specify the ontological commitments of the SR hypothesis such that it can be differentiated from the HO hypothesis. I close by suggesting that attempting to explain experience in terms of self-knowledge is misguided as self-knowledge and experience dissociate. At best the hypothesis requires more specific formulation as to the nature of self-knowledge.

Is the SR account of consciousness different from the HO account?

Kidd focuses on a particular experience in order to distinguish the SR account from the HO. This experience can be described as an epistemic intuition. Kidd considers two kinds of scepticism, scepticism about the external world and scepticism about one's own mental life. He suggests that whilst on reflection external world scepticism can appear to be a real worry, that we just aren't gripped by the possibility that our own minds don't exist, or even the more conservative possibility that one is not in a mental state that one seems to be in. It is the fact that we don't find this last possibility plausible which Kidd seeks to explain.

Such a feeling must arise, he supposes, given an SR account because it is in fact the case that being in a mental state is partially constitutive of it seeming to be in that state. We cannot seem to be in a mental state without actually being in that state. So it seeming that one is in a mental state justifies the belief that one is actually in that state. As such a kind of mental state scepticism doesn't appear plausible; one is always justified in believing one is in a mental state if one represents oneself as such.

But, he suggests, that HO accounts cannot guarantee that we will have this intuition. The reason for this is that being in a mental state is not, according to HO, partially constitutive of it seeming to the subject that they are in that state. He points to discussions of empty higher order thoughts as an example. On the HO account it is conceivable that a HO representation exists without the first order state that it represents also existing. A consequence of this is that it should be possible for one to seem to be in a mental state which one is not actually in. According to Kidd on the HO account it seeming to oneself that

one is in a mental state does not justify the belief that one actually is in that state. As such if HO were correct then we ought to find mental state scepticism plausible.

This example, however, doesn't seem to successfully distinguish the two accounts. The reason for this is that whether or not seeming to be in a particular mental state counts as a good justification for believing that one is in that state are *inaccessible* to the subject. Suppose for a moment that SR is true, there is still nothing in an experience which tells the subject that the experience is partially constituted by a mental state. Suppose I experience myself as seeing a computer screen in front of me. According to SR this is because I am in a first order mental (visual) state representing the screen and this partially constitutes the fact that I represent myself as being in such a state. But this relationship of partial constitution is not, so far as I can see, represented by the subject. Without this the justification for me believing that I see a screen is inaccessible to me and so cannot explain the experiential difference between external world scepticism and mental state scepticism. At least not until I adopt the SR hypothesis.

Similarly for the HO account. The relationship between it seeming that I see a screen and the first order state representing the screen is not represented to me, the subject. As such it is irrelevant to me that this relationship is not one of partial constitution. Whatever justification I have -or better, justification that I seem to have, we are talking about an experiential difference after all- for believing I see a screen it cannot be based facts about the relationship between (in Kidd's terminology) M and M*, because I do not represent those facts.

As such it seems that the SR and HO accounts cannot be differentiated by their explanations of particular epistemic feelings as neither has, of yet, offered a compelling explanation of those feelings.

What about some other standards in consciousness studies? Could the hypotheses be differentiated on, say, a binocular rivalry task? In such a task a different image, say a red car and a green house, is presented to each of the subjects eyes. The, somewhat surprising, outcome of doing this is that subjects don't see a mixture of the two images, rather they see and alternating sequence of images e.g. car, house, car, house. The HO account would begin to explain this by positing an alternating sequence of higher order representations. The first order visual representations of the car and house would remain (relatively) stable, whilst there would be an alternating sequence of HO representations representing which one of two visual states the subject is in. The SR account would begin the same way, but as we further develop the explanation the relationship between the higher order and first order states would be specified as one of partial constitution, rather than mere causation and representation.

The challenge to Kidd at this point then is to specify what difference this would make. I cannot see what difference this relationship between M and M* should make to experiences such as this. It appears to me that it is the representational relationship, rather than constitution or causation, between the higher order (M*) and first order (M) states which does all the explanatory work. What matters is that M* represents that I am in M, whether or not this relation obtains between distinct entities or within a single emergent entity doesn't seem to be doing any work.

Perhaps my arbitrarily chosen example is just one which happens not to be helpful? Sure the accounts make the same predictions for binocular rivalry, but it can't possibly follow from this that they make the same predictions for all conceivable tasks. This is true enough, however, I wish to return to the point above that the difference between the two hypotheses which Kidd discusses, namely what counts as justification for knowledge, is not accessible to the subject in guiding behaviour, nor does it clearly have *any* behavioural implications. This it seems to me arises from the fact that Kidd has focused on a normative epistemological difference, rather than an ontological one. The ontology of the two competing hypotheses regarding consciousness should therefore be further specified in order to allow us to distinguish the views.

The Epistemology versus the Ontology of Consciousness

As many readers will not doubt suspect, underlying the specific problems introduced above is a worry regarding the appropriate methodology for studying consciousness. In particular, regarding whether or not reflecting on the epistemology of self-representations can give us good insight into the ontology of consciousness. I would hate, however, to collapse into an argument about the "right" way to do philosophy. There is no single conclusion, or even set of consistent conclusions, to be drawn from such a debate[1]. Instead I wish to suggest that there is reason to doubt the tight link between self-representation/knowledge and phenomenal consciousness as required by Kidd's hypothesis.

I assume that Kidd does intend that all forms of self-knowledge necessarily lead to experiences. For instance, I know that I had knee surgery when I was 15, but thankfully this doesn't mean that I experience the surgery. The reason Kidd may give to this is that this knowledge is not (partially) constituted by a first order representation. There are other important differences too, the knowledge is retrospective and it is not knowledge of a mental state. I suspect then that the kind of self-knowledge needed to make something an experience must be i) knowledge that one-self (*qua* subject) is in a mental state, ii) that the

[1] That is not to say that I find methodological debates pointless (e.g. I think it's important to take seriously the criticism of a priori conceptual analysis that it assumes we have all of the requisite concepts need to understand something prior to empirical investigation), I just don't think entering into such a debate would achieve anything here.

mental state must be partially constitutive of the knowledge and iii) the relevant kind of constitution is synchronic (i.e. experiences are not historical entities, a la biological species).

This, I hope, is a fair characterisation of the kind of knowledge needed for experience according to Kidd; I am certain to be corrected if it is not. However, it seems to me that we can both change this kind of knowledge without changing an experience and change experience without changing this kind of knowledge.

Consider first an ordinary change blindness scenario. In these tasks subjects are shown an alternating series of two pictures. The pictures are largely the same except for one large difference that would be easily noticed where the pictures presented side by side. When one image changes to the next there is a flicker of distraction stimuli superimposed on the image.

You can see some examples here:
http://nivea.psycho.univ-paris5.fr/Mudsplash/Nature_Supp_Inf/Movies/Movie_List.html

Set up in this way it is very difficult for subjects to notice the change between the two images. It seems, at least initially, to the subject that they see the same image before and after the flicker. Nevertheless this can be accompanied by an epistemic feeling that something has changed, but that something is not identifiable. On Kidd's type of view this feeling it seems would be based on change in the content of the self-representation of the visual state. In his terminology the M* for the first image is different from the M* for the second. Further metacognitive monitoring of this gives rise to the epistemic feeling that something has changed. Nevertheless the experience of the two images is of them being the same.

If this is right then it seems that the right sort of self-knowledge for consciousness has changed from image 1 to image 2- giving rise to a feeling that something has changed, but there is no change in the visual appearance of the images. Image 1 still looks like image 2, even though, on watching the change, I have subjective knowledge that they are different. The challenge to Kidd then would be to offer an alternative account of the feeling of change that in no way depends on a change in M*, to explain why this kind of self-knowledge is different from the kind required for experience or convince us that image 1 and image 2 always appear different.

We can also examine the connection between self-knowledge of the M* type and experience from the opposite direction. There seem to be instances where subjects have the same self-knowledge for two different experiences. Consider a scenario in which you are a subject in a Sperling type experiment. On one trial you are presented with an array and the first row is cued. You successfully report the entire row. Well done! Now suppose on the

next trial you are presented with an almost identical array, except that the last letter on the third row is replaced with a non-alphanumeric character. On this trial the 2nd row is cued, and you again report it. Now in this scenario some more observant participants may claim that the same array was presented in both trials. So far as they know the experienced the same array twice, but they did not notice the single alteration (in fact subjects often do not notice non-alphanumeric characters on non-cued rows). Is it right to say that they did experience the same array each time? It is not obviously so- had the 3rd row be cued then the subject would have noticed the non-alphanumeric character on the second trial. Obviously I am being rather speculative here (and cue discussions on how to interpret this paradigm), but an open option here is that the subject had two different experiences but so far as they knew had the same experience twice.

Now this is not demonstrated, and I'm not sure that I've convinced myself that this scenario could really occur. However, all I want to do here is put some pressure on the idea that the right kind of self-knowledge is necessary and sufficient for phenomenal consciousness.

Conclusion

I summary then I have raised 3 challenges for Kidd in this comment. First, to show how experience can be sensitive to normative claims regarding epistemic justification thus allowing SR to be distinguished from HO. Second, to show how the proposed difference in relationship between M and M* can lead to a testable difference when it is the representational relationship which seems to be doing all the work and this is preserved between HO and SR. Finally, to show that there is the tight link between (a kind of) self-knowledge and experience as required by SR.