

# Self-Knowledge, Theoretical Knowledge and Science

Yakir Levin, Beer-Sheva

1. As part of his celebrated attack on the “Myth of the Given” Wilfrid Sellars argues for a radical alternative to the orthodox, neo-Cartesian conception of our knowledge of our own minds – *self-knowledge*, for short (Sellars 1997; unless otherwise indicated all references are to this work). The orthodox conception finds a particularly elaborate expression in traditional empiricism. On this conception, it is with our own mental states that we hold the most intimate and direct epistemic relation. Therefore, self-knowledge is a paradigm, indeed *the* paradigm, of non-inferential and non-theoretical knowledge. In contrast, Sellars claims that self-knowledge is akin to theoretical knowledge in science. Indeed, in his view self-knowledge illustrates the continuity of science with ordinary thinking (97). At the same time he takes self-knowledge to illustrate the fact that the distinction between the theoretical and the non-theoretical is merely methodological (84). These general points are made fairly clearly by Sellars. However, the argument that he offers on their behalf is rather obscure. The main purpose of this paper is to provide a clear reconstruction of Sellars' argument. But I shall also provide a brief diagnosis of a crucial weakness in the argument due to which it is, ultimately, unsuccessful.

2. Descartes, famously, considers awareness to be of necessity *discursive* i.e., concept involving: an awareness of something as something. In his view non-discursive beings that lack conceptual capacities must be mere automatons devoid of mental states (Kenny 1981, 243-5). In contrast, although their conception of the mind is in crucial respects Cartesian, traditional empiricists are committed to the possibility of a *non-discursive awareness*, i.e. one that does not involve concepts. Indeed, according to traditional empiricism discursive awareness rests on non-discursive awareness in two respects. First, episodes of non-discursive awareness are our basic means for forming and grasping concepts. Second, episodes of non-discursive awareness by which we acquire concepts warrant the application of these concepts in episodes of discursive awareness - viz. basic, non-inferential empirical beliefs - on which all other empirical justifications rest. I shall call the first thesis *non-discursive formationism*, and the second thesis *non-discursive foundationalism*.

A line of thought that has been taken to support non-discursive foundationalism is the following. If beliefs could be justified only by beliefs then either justification would be viciously circular or it would involve infinite regress. But the only episodes of discursive awareness that may justify beliefs are beliefs. Thus, for justification to be free of either vicious circularity or infinite regress it must rest in the last analysis on episodes of non-discursive awareness. This is particularly true of empirical justification in which case the most natural candidates for such justificatory episodes are non-discursive bits of experience. (Cf., Bonjour 1985, chap. 2 for much further elaboration of these considerations.)

Another line of thought that has been taken to support non-discursive foundationalism is the following (McDowell 1996, 3-7). Unless exercises of empirical concepts are rationally constrained from outside the conceptual sphere they cannot be considered as reaching beyond this sphere. They must be considered as “moves in a self contained game”, a mere “play of empty forms”, “a friction-

less spinning in a void”. The conceptual sphere may be thus constrained, it appears, only if non-discursive bits of experience constitute the ground level of empirical justification. But if so, non-discursive foundationalism is entailed by the very possibility of intentionality.

At this point the question arises of how episodes of non-discursive awareness can fulfill the role of “the tortoise on which stands the elephant on which rests the edifice of empirical [justification and] knowledge” (73). The answer suggested by traditional empiricism is based on non-discursive formationism. On this thesis, to be a bit more specific than above, classifications by way of non-discursive awareness of similarities and dissimilarities provide the basis for ostensive definitions that set up rules for the use of the terms applied in our empirical statements. According to traditional empiricism the basic beliefs on which all our other empirical beliefs supposedly rest are true “as a matter of ...following [these] rules for ... use” (72, 77). As such these beliefs resemble analytic statements, though the latter are true as a matter of following rules for use set up by linguistic definitions – i.e. definitions of linguistic expressions in terms of other linguistic expressions. So “in spite of the ... differences between [basic beliefs] and “analytic statements”, there is an essential similarity, [according to traditional empiricism,] between the ways in which they come by their authority” (71-2).

The considerations underlying non-discursive foundationalism explain in part the fundamental empiricist thesis that empirical justification and knowledge originate in experience. The account of the epistemic authority of experience in terms of non-discursive formationism explains in part another fundamental empiricist thesis – viz. that our concepts originate in experience. So the foregoing exposition has also unveiled a deep connection between these two fundamental theses.

Non-discursive formationism and non-discursive foundationalism are most familiar with respect to beliefs about the world around us. But they are also supposed to apply to beliefs about our own mental states. According to traditional empiricism these beliefs are basic beliefs grounded in episodes of non-discursive awareness. Moreover, traditional empiricism takes it that some beliefs about our own mental states – viz. beliefs about how things look or appear – are epistemically prior to beliefs about how things are in the world (32-46 passim). So the view that our access to our own mental states is the most immediate is part and parcel of traditional empiricism.

Non-discursive formationism and non-discursive foundationalism constitute the crux of what Sellars calls “The Myth of the Given”. His radical alternative to the orthodox conception of self-knowledge stems from his subtle and rich attack on these theses and their implications. It is to relevant aspects of this attack that I shall now turn.

3. Anything with stable dispositions to respond differentially to stimuli can be thought of as classifying the stimuli according to the repeatable responses those stimuli elicit. This is true of a thermostat that responds differentially to the temperature in its vicinity. And it is also true of the honeybee that responds differentially to the location and quality of the food it encounters on a foraging trip.

However, a classification by differential response is not yet a discursive classification by way of concepts. In order to count as discursive a classification must have three closely related features that take it far beyond a mere differential response. First, the classification must enjoy an appropriate distance, to use McDowell's term, from the direct causal influence of the items classified (McDowell 1996, 57). It must not be a mere stimulus-response process. To achieve such a distance the classification must, secondly, take up a position in "the logical space of reasons, of justifying and being able to justify" (40-1, 73-77). By being sensitive to the reasons relevant to a given classification the classifier may classify an item differently from how the direct causal influence of the item disposes her. Finally, the classification must involve language, indeed be a "linguistic affair" (62-64). For the inferential relations that characterize the logical space of reasons may obtain only between linguistic entities.

Classifications by mere differential response belong to the *non-normative* logical space of (causal) law. As such they cannot yield classifications that belong to the *normative* logical space of reasons. This means that for the classificatory processes assumed by the thesis of non-discursive formationism to yield concepts and language, these processes must involve concepts and therefore language. As Sellars puts it, "all awareness of sorts, resemblances, facts etc., in short, all awareness of abstract entities – indeed, all awareness even of particulars – is a linguistic affair" (63). But then the thesis of non-discursive formationism is incoherent. For its gist is that the concepts yielding processes it assumes are wholly pre-conceptual and pre-linguistic. Due to the close ties between non-discursive formationism and non-discursive foundationalism, the incoherence of the former thesis also reflects on the latter thesis. But I cannot go into this important point here.

4. The acquisition of concepts requires, then, a prior possession of concepts and language. This raises the question of how concepts can be acquired at all. Aware in part of this problem, traditional rationalists took it that at least some of our concepts, the most basic and general ones, are not acquired but innate. In the innateness debate, however, Sellars sides with the empiricists and assumes that *all* concepts are acquired (20-21). Given this assumption the processes by which we acquire concepts cannot involve concepts that we already own. These processes must therefore involve concepts that others own. They must consist of a training through "publicly reinforced responses to public [things] ... in public situations" by others who already possess concepts (86-7).

Obviously, mental concepts that are acquired by such a training must be "introduced in terms of a basic vocabulary pertaining to overt behavior" (100). This behaviorist conclusion – *methodological behaviorism*, as Sellars calls it (98) – may appear to involve a commitment to *analytical* or *philosophical behaviorism* – namely, the view that all proper mental concepts should be definable in terms of overt behavior. If this is indeed the case we would lose our grip on the idea of mental states as *inner episodes* – i.e. episodes that do not consist of overt behavior, and that are private in that each of us has a privileged access to her own. However, methodological behaviorism does not entail analytical behaviorism. This is so since "the behavioristic requirement that all concepts should be *introduced* in terms of a basic vocabulary pertaining to overt behavior is compatible with the idea that some behavioristic concepts are to be introduced as *theoretical* concepts" (100). In other words, even if all mental concepts should be introduced by reference to overt behavior, they need not

all refer directly to overt behavior. Some may thus refer only indirectly by referring directly to non-behavioral features that must be inferred in order to explain overt behavior.

It follows that the only way to retain the idea of mental states as inner, non-behavioral episodes is to consider these states as theoretical entities. Contra the positivistic conception of science, this does not mean that mental states qua inner episodes should be considered as of less authenticity than directly observable items, as useful fictions or as mere heuristic devices (83-4). But it implies a radical reorientation of the orthodox conception of self-knowledge, a reorientation that renders this type of knowledge akin to scientific knowledge. As such this reorientation illustrates the continuity of scientific and pre-scientific thinking. It illustrates, in other words, the fact that "the process of devising "theoretical" explanations of observable phenomena did not spring full-blown from the head of modern science" (96), but is "the flowering of a dimension of discourse which already exists in what historians call the 'prescientific stage' " (81).

5. If mental states are theoretical entities they must be intersubjective in that there is no difference between third and first person access to them. But doesn't this mean that if mental states are theoretical entities we lose our grip on the idea that they are private in that each of us has a privileged access to her own? If "privacy" here means "absolute privacy" – i.e. a privileged access that is independent of context – then the intersubjectivity of our mental states does indeed entail that they are not private (107). However, there is a weaker sense of privacy, which is compatible with intersubjectivity; indeed, privacy in this sense presupposes intersubjectivity. Thus, people who have been taught a theory that applies to their behavior "can be trained to give reasonably reliable self-descriptions, using the language of the theory, without having to observe [their own] overt behavior. [This may be brought] about, roughly, by applauding utterances by [the trainee] of [e.g.] 'I am thinking that p' when the behavioral evidence strongly supports the theoretical statement '[The trainee] is thinking that p'; and by frowning on utterances of [e.g.] 'I am thinking that p' when the evidence does not support this theoretical statement" (106-7). But once one has been trained in this way one has gained a sort of privileged access to one's mental states. One may then reliably report on one's mental states without relying on any behavioral evidence, while others cannot do this. "What began as a language with a purely theoretical use has gained a reporting role" (107).

This illustrates a general point about theoretical concepts (79-85; Brandom 1997, 162-166). According to Sellars purely theoretical concepts are ones we can be entitled to apply only as the conclusions of inferences, whereas concepts of observables also have non-inferential, reporting uses. Given these definitions, the line between the theoretical and the observable may shift with time or with the right training. Thus, Pluto was introduced as a purely theoretical object. But the development of more powerful telescopes eventually made it a subject of non-inferential reports. Similarly, physicists with the right training can non-inferentially report the presence of mu mesons in bubble chambers. So "the distinction between theoretical and non-theoretical discourse [is methodological rather than] substantive" (84) .

The notion of observation underlying the orthodox account of self-knowledge is that of direct perceptual or introspective acquaintance. Obviously, this notion is very different from the notion of observation underlying Sellars'

conception of self-knowledge – viz. non-inferential reporting, which is made possible by a training akin to conditioning. In particular, unlike observational concepts in the former, orthodox sense, those in the latter, Sellarsian sense “are primarily and essentially *intersubjective*, ... and [their] reporting role ... constitutes a dimension of [their] use which is *built on* and *presupposes* this intersubjective status” (107). So although Sellars takes mental concepts to shift status from the theoretical to the observational, this does not narrow the wide gulf between his account of self-knowledge and the orthodox account.

6. Sellars' conception of self-knowledge is principally based, then, on a rejection of a fundamental assumption of the empiricist conception – namely, that concept acquisition is based on non-discursive and non-linguistic processes. As against this assumption Sellars argues that the acquisition of concepts must be a discursive-cum-linguistic affair. From this he draws the conclusion that concept acquisition must be a communal affair – i.e. that we can only acquire concepts by way of training by other subjects of awareness who already possess concepts. Mental concepts acquired in this way must refer first and foremost to overt behavior. So to avoid philosophical behaviorism Sellars concludes that our concepts of mental states qua inner episodes distinct from overt behavior must be theoretical concepts that refer to overt behavior indirectly.

Sellars' crucial step in this ingenious argument - his claim that concept acquisition must be a discursive-cum-linguistic affair - seems correct, or at least I shall grant him this claim. All this claim entails, however, is that the

acquisition of concepts must involve the prior possession of *some* concepts, perhaps the most basic and most general ones. On this basis Sellars may conclude *at most* that the acquisition of the latter concepts must be a communal affair. But this conclusion is compatible with the possibility that the acquisition of all other concepts is *not* a communal affair. In particular, insofar as Sellars' argument is concerned the acquisition of basic concepts such as that of property, identity, similarity etc. by way of communal training may enable one to construct mental concepts in a non-communal manner – e.g. by way of introspecting similarities between one's inner episodes in a *discursive* manner that involves the basic concepts. This very brief diagnosis of a crucial weakness in Sellars' argument requires further elaboration. But it suffices to show, I believe, that this argument is, ultimately, unsuccessful. The demise of the Myth of the Given does not entail that self-knowledge is akin to scientific knowledge.

## Literature

- Bonjour, L. 1985. *The Structure of Empirical Knowledge*, Cambridge, MA: Harvard UP.
- Brandom, R. 1997. “A Study Guide of *Empiricism and the Philosophy of Mind*”, in Sellars (1997), 119-181.
- Kenny, A. (ed.) 1981. *Descartes: Philosophical Letters*, Oxford: Basil Blackwell.
- McDowell, J. 1996. *Mind and World*, Cambridge, MA: Harvard UP.
- Sellars, W. 1997. *Empiricism and the Philosophy of Mind*, Cambridge, MA: Harvard UP.