Self-Knowledge, Introspection and Memory

Self-knowledge (S-K) has a variety of dimensions and philosophers have proposed, and continue to propose, a host of different methods to explore or to establish the principles of a very particular form of knowledge such as this. What is quite impressive in the philosophical literature (by which I mean not only that of analytical inspiration but also that of the so-called continental philosophy) is the fact that the usual relationship between S-K, introspection and reflection leaves no space for the essential role of memory. Hence, it is memory’s role in S-K that I wish to look at here.

Traditionally, empiricist and rationalist approaches have dominated the epistemic views of S-K. The latter starts with Descartes, or even earlier, and continues until Kant; the former includes Locke and Hume. In fact, if one thinks of Hume and Kant as the most relevant modern representatives and the culmination points of both approaches to S-K, it would seem perfectly acceptable to suggest that introspection plays a determinant methodological role. In the first case, let us remember what Hume famously said regarding his own mind and what he finds there: "For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of hot or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself, at any time without a perception, and never can observe any thing but the perception" (D. Hume, 1984, 300). In the case of Kant, in the section

---

1 A first draft of this paper was presented at the seminar about “Knowledge and Ability” coordinated by Franck Lihoreau. I’m grateful for his kind invitation and for the stimulating discussion with the researchers who participated in that session.

2 This is a too simple if not erroneous scheme taking into account the multiple and often contradictory positions that exist even inside the same field. It is enough to stress that Descartes (or what can be designated as Cartesianism), who is pointed out as the classical paradigm of self-knowledge by introspection (S-KI) and the concept of reflection that operates in it, is submitted to a devastating critique in Kant’s first Critique (notably in the “Paralogisms of Pure Reason” of the first Critique). I do not think there is enough evidence in Descartes’ main works to support the picture referred to above, but I shall not follow up this discussion. The fact is that it is in Locke that a clear concept of reflection as an essential piece of introspection and self-knowledge is to be found. The following quotation from Locke’s Essay Concerning Human Understanding illustrates a relatively common concept of reflection which prevails in the literature on these themes: “The mind receiving the ideas mentioned in the foregoing chapters from without, when it turns its view inward upon itself, and observes its own actions about those ideas it has, takes from thence other ideas, which are as capable to be the objects of its contemplation as any of those it received from foreign things” (Essay ..., II, chapter VI). Objects “as capable to be the objects of its contemplation as any of those it received from foreign things” are precisely what seem to be at the core of any introspective move, but this is precisely what makes it so problematic.
"The paralogisms of pure reason", because he is occupied with the deconstruction of all dogmatic metaphysics about the 'I', the introspective tool is also used: "Now in inner intuition there is nothing permanent, for the 'I' is merely the consciousness of my thought. So long, therefore, as we do not go beyond mere thinking, we are without the necessary condition for applying the concept of substance, that is, of a self-subsistent subject, to the self as a thinking being. And with the objective reality of the concept of substance, the allied concept of simplicity likewise vanishes..." (Kant, Critique of Pure Reason, B 413, transl., N. Kemp Smith).

With these quotations from these two classical authors I only wish to show that knowledge of the self, which should be acquired by introspection, is relatively poor: a particular kind of perception in the case of Hume, and the mere consciousness of the ‘I’ in the case of Kant. I make this remark in order to show that scepticism about the possibility of genuine knowledge already has, to my mind, its roots in classical philosophy, be it empiricist or rationalist. When I refer here to scepticism in relation to the knowledge of one’s own mind (or self), I mean our capacity to know our mind by setting up a dual structure, that is, a mind that looks at itself, a mind that sets itself up as a target. Surprisingly, if one takes the examples of both Hume and Kant (for different reasons), this target seems to offer us little: a perception of feelings or awareness of an ‘I’.

Nonetheless, in order to approach the notion of self-knowledge, one needs to understand as clearly as possible the link between reflection and memory, and my argument below is based on the view that the dual structure of reflection has to be revised by the introduction of the role of memory. Consequently, what I suggest is that clarification of what can be considered self-knowledge and introspection requires a consideration of the role that is played by memory and the evaluation of the epistemic nature of retrodictions (that is, statements about past events, in this case, statements about past mental events). Then my argument, which for the moment is only designed to be a draft, goes through the following points:

---

3 In the same vein, Wittgenstein quoting William James makes a nice (and intentionally oversimplified) picture of introspective operations: “Here we have a case of introspection, not unlike that which gave William James the idea that the ‘self’ consisted mainly of ‘peculiar motions in the head and between the head and the throat’. And James’ introspection showed, not the meaning of the Word ‘self’ (so far as it means something like ‘person’, ‘human being’, ‘he himself’, ‘I myself’), or any analysis of such a being, but the state of a philosopher’s attention when he says the Word ‘self’ to himself and tries to analyse its meaning. (And much could be learned from this.)” (Wittgenstein, PI 413).
- I: Introspection is an inescapable tool of S-K and is an instance of direct knowledge of my mental or psychological/mental states, then S-K is in most cases S-K by introspection (S-KI).

- II: S-KI is memory dependent (there is a structural time factor in all S-KI that is virtually unrecognized by many philosophers).

- III: if S-KI is dependent on memory and at the same time is direct (unmediated) knowledge, it cannot be treated as an analogy with displaced perceptual knowledge (displaced knowledge: for example, to know that person A was in this room by seeing his hat).

- IV: S-KI is expressed in retrodictions, that is, statements about my own past mental states (not to be confounded with statements about myself in the past).

- V: S-K that is not S-KI does not depend on memory. Typically, it corresponds linguistically to expressions or avowals in the present tense which possess a different epistemic nature, namely they are not only authoritative in the first person but are also infallible. Typically, expressions like “I’m cold”, “I’m furious”, “I’ve the intention to do this”, etc., which are expressions of mental/psychological states, require only sincerity conditions. Yet I shall suggest that even expressions or avowals are not totally memory independent if one accepts the working memory function introduced by some psycholinguists. But this issue is not to be discussed here.

- VI: S-KI does not have a reflexive structure, in the sense that it is not a second-order thought targeting a first-order one.

- VII: S-KI has first-person authority with respect to the contents of one’s own mind although it is fallible precisely because the retrodictions are linguistically memory dependent.
VIII: Statements like “I was sad in Paris last week”, which possess first-person authority, are fallible and are pieces of S-KI unlike the correspondent statement “I was in Paris last week”, which does not have any authority of that kind and is consequently not S-KI. Nevertheless, let us mention that both statements are fallible.

Below I will not deal with each of the above points separately; I only wish to identify some transversal guidelines against the background of the views of some of main authors. Let us begin with Fred Dretske, whose book *Naturalizing the Mind* (1995) has an entire chapter entitled “Introspection”. An intuitive definition of it would be, for example, “the mind’s direct knowledge of itself” (Dretske, 1997, p. 39) although he adds that introspection is an “instance of displaced knowledge”.

“Introspective knowledge is knowledge of the mind – i.e., mental facts. Since mental facts (according to the thesis I am promoting) are representational facts, introspective knowledge is a (conceptual) representation of a representation – of the fact that something (else) is a representation or has a certain representational content. It is, in this sense, *meta-representational*. Meta-representations are not merely representations of representations. They are representations of them *as* representations” (1997, 43). “Since mental facts (according to the thesis I am promoting) are representational facts, introspective knowledge is a (conceptual) representation of a representation – of the fact that something (else) is a representation or has a certain representational content” (ibid.)

Then the examples Dretske gives are analogies with representations of external objects: I can see a photo *as* a piece of paper, *as* an object with 2 grams, *as* a photo of Clyde, and so on. I suggest that this feature of introspection as representation of one’s own mental states *as* something else seems quite hard to accept. Normally I do not represent me *as* sad, but I am or I was or I will be simply sad. Cases in which I represent me *as* sad are typically situations where I say that I represent my face *as* sad (either in a picture or in a photo or in my imagination). But these situations are not pieces of S-K.

Of course a little more introspective attention on my recalling can raise doubts in me about my sadness and then I will ask myself whether what I felt was a genuine
feeling of sadness, about the reason I was sad, and so on. Anyway in such cases it would be better to speak about continuous and associated memory operations. However, these cases do not have the double form of a first-order representation, which is set as a target, and another second-order one, that is, a meta-representation of that target. Even a retrodiction of some complexity such as “I was sad in Paris because of this and that, and so on” does not have a dual structure like a meta-representation of first-level representations. It is much more analogous to my direct representation of myself in a mirror. The situation that just simulates the meta-representation would be: I am seeing myself in a second mirror, which mirrors me in a first mirror. But this does not adequately simulate S-KI: remembering my past mental state, when I was sad, does not have any kind of intermediate representation; it is direct knowledge (occurred at a time 2) that essentially consists of the recalling of a past event (occurred at time 1). Recognizing this difficulty Dretske claims that:

“It may seem as though this account of introspective knowledge – as a species of displaced perception – makes it into a form of inferential, and thus indirect knowledge. If introspective knowledge of oneself – that one represents the world thus and so – has the same structure as knowing that the postman has arrived by hearing the dog bark, then there is an intermediate ‘step’ in the reasoning that makes knowledge of the target indirect. There are two ways of knowing that the postman has arrived: by seeing or hearing him arrive and by seeing or hearing something else (the dog) that ‘tells’ you he has arrived. If, on a representational theory, introspective knowledge is more like the latter than the former, then the representational account fails to give self-knowledge the immediacy that we know it to have” (Dretske, 1997, 60)

Then Dretske adds that this is a relevant objection to the analogy that he claims between introspection and displaced perceptual knowledge, but he insists that in an introspective statement like “I represent k as blue” the introspective quality comes (and this is the fundamental difference between displaced introspective knowledge and displaced perceptual knowledge) from the fact that the truth of the information of my representation of blue is not relevant, if what I see/represent is or is not blue; the relevant fact in that introspective statement is about I am representing k
To my mind Dretske wishes to preserve the traditional dualistic structure of introspection and self-knowledge as reflection.

This leads us to another point about the reflexive nature of introspection and S-K. Like Dretske, T. Burge, or more recently P. Jacob, accepts this reflexive structure as part of S-K. I have already commented on Dretske’s view of displaced knowledge, on his wish to preserve the reflexive feature of introspection and the implicit difficulties he recognized. In the case of Burge, it is possible to find difficulties regarding S-K as a dual reflexive structure although of another kind to those that are to be found in Dretske. The case of Tyler Burge is quite specific in the sense that he considers that, on the one hand, at least some S-K statements seem not to require a dual structure since the targeted thought coincides in time and content with the subject (second-order) thought, but on the other hand S-K involves necessarily the difference between meta- and first-level thoughts on which is based all reflection. Let us observe in some detail Burge's approach to this topic.

Burge sees the differences between S-K and perceptual knowledge in the personal status of S-K. Namely in S-K acts, one is facing statements that are made simultaneously from and about one first-person point of view. He identifies this feature in the following terms: "the point of view and time of the judgment must be the same as that of the thought being judged to occur" (Burge, 1988, 651). And he adds "When I judge: I am thinking that writing requires concentration, the time of the judgment and that of the thought being judged are the same" (ibid.). This also means, as Burge makes clear in another passage of the same paper, that "The reflexive judgment simply inherits the content of the first-order thought" (ibid., 650). The essential difference in relation to perceptual knowledge consists of the personal status of S-K, or in other words the use of first-person pronouns, and as he notes: “In all cases of authoritative self-knowledge, even in those cases which are not ‘basic’ in our sense, it is clear that their first-person character is fundamental to their epistemic status” (ibid. 651). Thus it is first-person authority and not so much its reflexive structure that confers a special epistemic status on S-K if compared with perception. Or perhaps in other words, the correct interpretation of Burge’s view is that it is the

---

4 For a detailed discussion of Dretske’s model of displaced knowledge applied to S-K, see P. Jacob, 2010. The theoretical framework of the discussion is Jacob’s wish to show that “our picture of introspective self-knowledge derives from two traditional sources: rationalist epistemology and empiricist epistemology” (ibid., p. 3).
reflexive structure of S-K plus the first-person authority that gives it its peculiar epistemic nature. Yet if one accepts Burge’s description of S-K, when one compares the epistemic nature of S-K and that of perceptual knowledge, one feature of the former (the thought I have has the same content and coincides in time with what is thought about) does not place both knowledge types far from each other. After all, in reflexive knowledge (as all S-KI should be) more than one content is not to be found. What makes the difference is the place of first-person authority in S-K. Yet does this difference produce drastic epistemic consequences? In Burge’s view it does, but I think it is a mistake: first-person authority does not imply infallibility in the process of introspection. Let us make the nature of this process clearer.

Using such criteria, it seems hard to maintain that S-K is clearly distinct from the usual perceptual knowledge since, after all, in both there is one content and there is no time gap that I can be aware of. My concern lies not so much in the claim for the existence of a unique content in the case of reflexive knowledge (“reflexive judgment simply inherits the content of the first-order thought”), with which I agree, but much more with the claim that in S-K there is a typical reflexive structure with a dual composition based on a targeted object by a second-level thought/representation. As I have tried to show, the reflexive operation does exist in all S-K yet as an introspective move which is memory dependent. S-K is S-KI with a dual time structure where memory is necessarily involved. By dual time structure I mean the awareness that a past t1 event is targeted by an actual thought/representation typically expressed in retrodictions at t2. This is exactly why Burge does not see in basic S-K a necessary time gap, which requires the work of memory. As I said above, he sees S-K as

---

5 This is a remarkable result in order to re-design the traditional concept of reflection and introspection. It is interesting to consider another feature of this redesign since in S-K each targeted event can be seen as a “recognizing” of what happens or of what is happening in me. Wittgenstein’s approach to the nature of recognition is striking inasmuch as it describes the misunderstandings surrounding recognition when memory is taken into account. The result is also “fusion” into one content, as is the case in Burge’s approach. “It is easy to misconceive what is called “recognizing”; as if recognizing always consisted in comparing two impressions with one another. It is as if I carried a picture of an object with me and used it to identify an object as the one represented by the picture. Our memory seems to us to be the agent of such a comparison, by preserving a picture of what has been seen before, or by allowing us to look into the past (as if down a spyglass)” (PI 604). Yet this is a mistaken picture, “Indeed, it is not so much as if I were comparing the object with a picture set beside it, but as if the object coincided with the picture. So I see only one thing, not two” (PI 605).

6 The fact that in S-K there is one content (first-order thoughts do not have any content on their own) does not imply that the targeted content that occurred in the past coincides in time with the actual statement/thought. On the contrary, it is essential to S-K that the awareness of a time gap (even the narrowest) is preserved. Yet if I am not wrong, Burge’s position erases any time gap whatsoever.
possessing the classical dual structure plus first-person authority and it is this characteristic alone that gives S-K a special epistemic nature. I would say that he means something near to infallibility. The following lines illustrate, in my opinion, this conviction: “The source of our strong epistemic right, our justification, in our basic self-knowledge is not that we know a lot about each thought we know we have. It is that we can explicate its nature and its enabling conditions. It is that we are in the position of thinking these thoughts in the second-order, self-verifying way” (Burge, 1988, 654).

But this is already problematic for Davidson to whom “of course people have beliefs, wishes, doubts, and so forth; but to allow this is not to suggest that beliefs, wishes, and doubts are entities in or before the mind, or that being in such states requires there to be corresponding mental objects” (D. Davidson, 2001, 35-6). So if there are no mental states as objects in the mind that one could grasp in any sense, there is no such a thing as introspection in the usual sense for Davidson I guess. But Davidson’s point is that such a thing is not needed in order to have an attitude. In Davidson’s words, “having an attitude is not having an entity before the mind; for compelling psychological and epistemological reasons we should deny that there are objects of the mind” (ibidem, p. 36). I believe Davidson has a fundamental point here, but I think he could go deeper in the consideration of these “compelling psychological and epistemological reasons”. It is at this moment that the problem we face can be formulated in the following terms: it is counter-intuitive to simply deny the existence of introspection and the correspondent S-K, but one should not count on dual reflexive structures, on target objects in the mind and similar sorts of objects to identify introspective knowledge. It is at this point that it seems to me necessary to introduce the time factor, or in other words memory. S-KI needs a time distance, even a minimal one: the statement about my own mental/psychological state made at t2 always refers to a t1 event. There is, I agree, a dual structure, but it is a present/past dual one not a reflexive one. Even I “reflect” at t2, for example, as a result of mental states that I felt at t1. There is always a dual time structure that still requires at least two things that are usually required for S-KI, which are 1) to be direct, that is, not mediated like any displaced knowledge, and 2) to have first-person authority.
Recent studies on memory and especially short-term memory elucidate us not only about the complex neuro-chemical processes involved and what distinguishes this kind of memory from long-term memory. They also allow us to understand the role that short-term memory plays as it works when representations/thoughts coincide (in time) with their correspondent content ones or when we report something to other people or to ourselves. Psychologists call this very short-term memory “working memory” and it can be defined in the following terms: “Working memory can be thought of as a low capacity information reservoir that is always full, sensations flowing into it continuously at about the same rate that they are forgotten (…). Working memory is an indispensable form of transient memory; it is a moving window of comprehension that allows us to understand the present in terms of the very recent past” (O’Shea, M., 2005, 85). This is the case whenever the grammatical form is equivalent to present tense statements. It is enough to think about statements like “I’m understanding this text”, “I’m telling you how much I am sad”, “I’m seeing this sunset as a Rothko picture”, etc. These are statements that do not seem to have the same structure as simple avowals or expressions. Precisely they involve a working memory that enables cognitive performances and are present in all language learning.

This way of seeing S-KI leads us to the question of the status of the content of retrodictions, which are memory dependent statements and are different from retrodictions that are not memory dependent, or rather that have a truth-value that does not depend on my memory: the statement that I was in Paris last week is a retrodiction of this type. It is not a piece of introspection, and even if it possesses the dual time structure, it lacks the other essential component, that is, the authority of first-person.

---

7 For many psycholinguists working memory is implicated in syntactic comprehension or in keeping track of syntactic dependences. Furthermore, the working memory is what enables cognitive performances such as attention being focused on targets at a particular time. For example, the judgment, “Now I’m understanding this argument” involves working memory, and even implicit elements learned in the past that enable comprehension are, so to say, linguistically hidden. If we consider that judgment is a piece of S-K, then the targeted element is memory dependent and the statement itself possesses a dual time structure.


8 First-person authority must fulfil the following condition: a total asymmetry in relation to the truth/satisfaction/conditions of the correspondent third-person statement’s content. Then “I (A.M.) was in Paris last week” has the same truth/satisfaction condition of “He (A.M.) was in Paris last week”.

My past sadness is not a target of a meta-representation but simply the content of a thought/representation that in this case is a recall of a past state of mind, experience or whatever. What happens is this: the so-called first level content, i.e. the event that at time1 I was sad, is recalled at time2. Is the t2 representation a meta-representation of the t1 representation? I would say that it is not and I shall illustrate this claim with a remark of Wittgenstein’s from the Philosophical Investigations.

“Does it make sense to ask ‘How do you know that you believe that?’ – and is the answer: ‘I find it out by introspection’? In some cases it will be possible to say some such thing, in most not. It makes sense to ask, ‘Do you really love her, or am I only fooling myself?’, and the process of introspection is the calling up of memories, of imagined possible situations, and of the feelings that one would have if…” (PI 587).

The remarkable thing in this quotation is that for Wittgenstein introspection, so to speak, achieves its target in most cases by an operation of recalling, of memory. In other words, in most cases the structure of introspection requires a time factor, and without it there is no target whatsoever. (In this case, the target element is the quality of a feeling.) On the other hand, when memory does not play a role, it seems that Wittgenstein would reject introspection. Without memory, no target, no introspection. Possibly the cases where Wittgenstein would say that it does not make sense to speak of introspection are such as: believing p and observing my belief seems to be absurd. Let us note that even some statements like “Now I’m believing p” implicate that I perhaps did not believe p before. But this contrast only needs the time factor. In fact, what is more plausible according to the quotation above is that Wittgenstein considered that in some cases where time did not play a role there was no sense in the question. Is memory dependent inquiry truly self-knowledge and are other forms, which exclude memory, pseudo self-knowledge? Wittgenstein does not make this explicit, but one can suspect that it is what he means.

Thus, as a non-definitive conclusion, S-KI has at least 3 features: a dual time structure based on memory, it must be direct and it must have first-person authority.

---

They are not asymmetric. But “I (A.M.) was sad last week in Paris” has not the same truth/satisfaction condition of “He (A.M.) was sad last week in Paris”. So “I was sad last week in Paris” has a first-person authority which is lacking in the first statement “I was in Paris last week”.
References:


Jacob, Pierre (2010), *Do we know how we know our own minds yet?* URL= http://jeannicod.ccsd.cnrs.fr/docs/00/05/34/50/HTML/index.html


Locke, John (1690), *Essay Concerning Human Understanding*


Wittgenstein, Ludwig (2009), *Philosophical Investigations*, (revised 4th edition), Oxford: Wiley Blackwell (quotations from *PI* are made in the usual form referring to the number of the section in the case of the first part of the work)