****Forthcoming in Avant: Special Issue on Thinking With Images****

Memory, imagery, and self-knowledge

ABSTRACT

One distinct interest in self-knowledge is an interest in whether one can know about one's own mental states and processes, how much, and by what methods. One broad distinction is between accounts that centrally claim that we look inward for self-knowledge (*introspective methods*) and those that claim that we look outward for self-knowledge (*transparency methods*). It is here argued that neither method is sufficient, and that we see this as soon as we move beyond questions about knowledge of one's beliefs, focusing instead on how one distinguishes, for oneself, one's veridical visual memories from mere (non-veridical) visual images. Given the robust psychological and phenomenal similarities between episodic memories and mere imagery, the following is a genuine question that one might pose to oneself: "Do I actually remember that happening, or am I just imagining it?" After critical analysis of the application of the transparency method (advocated by Byrne 2010, following Evans 1982) to this latter epistemological question, a brief sketch is offered of a more holistic and inferential method for acquisition of broader self-knowledge (broadly following the interpretive sensory-access account of Carruthers 2011). In a slogan, knowing more of the mind requires using more of the mind.

> [W]hat must become of all our particular perceptions upon this hypothesis? All these are different, and distinguishable, and separable from each other, and may be separately considered, and may exist separately, and have no need of any thing to support their existence. After what manner therefore do they belong to self, and how are they connected with it? For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat 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, *Treatise* 1.4.6

The above passage is familiar to any philosopher. Indeed, most of us have been compelled by it at one time or other in our careers. Centrally, Hume uses this observation to motivate his exorcism of the self: at any one moment, the most that introspection will reveal is some occurrent mental states. The idea that one may have of a substantial self, standardly promoted by philosophers up to Hume's day, is thus not grounded in experience, but instead is a fiction created by the imagination. If 'self-knowledge' refers to some kind of certainty about this imagined 'I', then Hume's conclusion is negative: there is no such thing to know. But 'self-knowledge' can also denote, and perhaps more commonly does in contemporary philosophy, an awareness or certainty of one's own mental states. And here Hume seems to indicate that this is a kind of knowledge that is readily available. One can distinguish both the way things appear to one—hot or cold, light or shade—and the character or category of state(s) that one currently enjoys—love, hatred, pain, pleasure.

A distinct interest in self-knowledge concerns whether one can know about one's own mental states and processes, how much, and by what methods. Hume's method, perhaps surprisingly, is introspection, and he gives clear indication that this method is a good one, enabling certain awareness of both the content and the character or category of one's mental states. Some contemporary theorists follow Hume at least this far: they take self-knowledge to both enjoy a

special kind of epistemic security and claim that this security partly depends upon the first-person privileged access that a subject enjoys, as it were, to herself. But from here there are a variety of accounts of the methods by which we do or should acquire self-knowledge. One broad distinction is between accounts that centrally claim that we *look inward* for self-knowledge; we use introspection. Opposite this kind of account, some claim that we *look outward* for self-knowledge, through the transparent mental states that are the targets of our inquiry. It will be argued here that neither method is sufficient, and that we see this as soon as we move beyond knowledge of one's beliefs.

The dominant emphasis in the contemporary self-knowledge literature is on knowledge of doxastic commitment, particularly belief. This is not surprising, since belief still plays a central role in epistemology, theories of rationality, and practical reasoning. How one might come to certainty about one's beliefs is therefore of obvious importance. But this is not exhaustive: the human mind involves a great deal more than belief states – intentions, memories, imaginings, goals, values, as well as sensory perceptual states, pains, itches, and so on. These states or processes also figure largely in decision-making, planning, action, well-being, and so here too certainty about one's mental states, more broadly and including these kinds, is of obvious importance. Much less work has been done, however, on self-knowledge broadened in this way.

One clear exception is recent work by Alex Byrne, who writes "that a familiar 'transparent' account of knowledge of one's beliefs can and must be extended to mental states in general" (2011: 202). Byrne argues that the transparency account can be extended to account for knowledge of one's perceptual experiences and to demarcate, for oneself, one's memories from one's mere mental images (Byrne 2010). Byrne's approach is admirable, since it is clearly sensitive to the varieties of self-knowledge, eschewing the myopic emphasis on knowledge of belief. But while broadening the scope of a theory of self-knowledge in this way is needed, the pure transparency method fails in this broader explanatory project, or so it will be argued here.

The analysis proceeds as follows. Section I offers brief clarification of the relevant notions of memory and imagery, plus a pair of questions about these phenomena. These are the central explananda for Byrne (2010) and provide a centrepiece for the critical and positive analyses given in the present paper. Section II briefly clarifies answers (including Byrne's) to the first, ontological question: what distinguishes memory from imagery as mental kinds?[1] Section III focuses on the second, epistemological question: how can one determine for oneself whether one is remembering some event or merely visually imagining it? Byrne's answer to this question extends some remarks made by Gareth Evans (1982). But this method—the *transparency method*—fails to do the work for which Byrne enlists it. More broadly (section IV), such a "looking outward" method fails to secure self-knowledge when one aims to discover not just the content of one's mental state, but also what kind or character of mental state one is in. Again, moving beyond belief here is important: an analysis of self-demarcation of memory vs. imagery reveals that the method proposed by transparency accounts is insufficient. But these accounts are not alone in this failure: pure introspective or "looking inward" accounts similarly fail to provide broadened self-knowledge of both the content and character of one's varied mental life. This result might encourage some to conclude with a familiar philosophical scepticism. But the more optimistic conclusion is that while (mental) self-knowledge might not come as easily as Hume and others have suggested, it is acquirable but requires a more complex or holistic method for its acquisition. The paper closes with some brief suggestions on how such an account would look.

[1] Numerous terms are used for distinguishing mental states as kinds: 'categories', 'characters', 'modes', 'attitudes', as well as 'kinds' or 'types'. There may be important differences in the meanings of some of these terms, and they are certainly not used univocally across the literature, but these complications will not be addressed here. The terms 'kind' and 'character' will typically be used, where canonical (even if controversial) examples of distinct mental kinds are beliefs, desires, intentions, memories, imaginings, visual experience, auditory experience, haptic experience, visual images, auditory images, and so on.

I. Imagery and memory

Like most mental terms, 'memory' is used to refer to a variety of phenomena. The relevant phenomenon for this discussion is *episodic memory*. One can recall, visually, events that occurred in one's life and when one does so, one enjoys a mental episode that is representationally and phenomenally similar to the experience remembered.[2] One might remember when one first met one's partner, or spotting a bear in the woods, or the emotions on the faces at a funeral. So episodic memory is first personal, in the rich sense that one remembers being or doing or experiencing at some previous time in one's own life, and to remember this is partly to remember, as we say, being there. (By contrast, semantic memory involves recalling, in a third personal way, that some proposition is true or that some event happened.) Episodic memory on this understanding is *factive*: if one remembers having an experience *E*, then it is true that one had *E*. Some take this to be a conceptual point about memory: all memories are ways of knowing, or ways of truly representing past events.[3] A weaker commitment suffices here: one kind of memory—episodic memory—is factive in this way.

Imagery also may take many forms, and there is no agreed upon definition for the phenomenon. The following characterization suffices for present purposes. First, mental imagery is a conscious experience that resembles the representational structure and phenomenal character of perceptual experience. The most familiar way to characterize this is as specific to a particular modality. Thus, when one visually images, one represents features typically perceivable through vision——colours, shapes, motion, depth——and as bound into an object. Consequently, what it is

[2] The emphasis here is on visual episodic memory and visual imagery, but this is not to suggest that there are no versions of either that are in, as it were, other modalities. Both audition and touch would seem to be good candidates. And moreover, it is plausible that much of our memory experiences and our imagery experiences are multi-modal, at least in the sense that they involve a stream of images in multiple sensory modalities, accordingly varying in their representational structure and phenomenal character. In the case of memory, when one remembers "being there" one does not just image the looks or the sounds or one's bodily position in space but, often, all of this and more.
[3] Williamson 2000; Cassam 2007.

like to visually image one's dog lounging in the afternoon sunlight is phenomenally similar to what it is like to visually perceive one's dog lounging in the afternoon sunlight. Likewise for imagery in any other sensory modality (or in multiple modalities, supposing multi-modal imagery is possible). But of course, imagery is not perception (or so it will be assumed here). And one way in which the two experiences are distinct is this: imagery, by contrast to perception, is typically not had in the presence of the or an appropriate stimulating cause(s). Away at a conference and feeling homesick, I visually imagine my dog lounging in the afternoon sunlight. I *could* do this in the presence of my dog, but it is rare that I would. And, by contrast, except in extremely rare cases (cases dreamt up only by philosophers, some might say), I do not visually experience my dog lounging in the afternoon sunlight unless I am in the physical presence of my dog in that very environmental context. Conjoining these two observations: A visual image is a quasi-perceptual mental state or process that (i) resembles the representational content and phenomenology of a visual perceptual experience; (ii) is typically had in the absence of the appropriate external cause. This is a working characterization not a definition.[4]

With these two phenomena in mind, as characterized, a number of similarities between episodic memory and imagery come into view. The two phenomena are typically both conscious experiences, and are both available for higher-level cognition or use in reasoning. Focusing again on the visual cases: memory and imagery are similar both with respect to representational structure and visual phenomenology. And, perhaps additionally, both are or can be perspectival. So, you might visually remember being present for Obama's first inaugural speech, while I merely visually image that speech (not having been present). These two experiences may be subjectively very similar with

[4] Additional distinguishing features that might complete a definition include the following. Hume characterizes images (by contrast to perception) as less vivid and immediately voluntary. And imagery plausibly relies on previous sensory information, while perception does not. More recently, some have argued that the determinacy of perceptual content is given by the world while the determinacy of imagistic content is given by one's memories (Nanay, 2015); or that imagery lacks the assertoric force that perception enjoys (Stokes and Biggs 2015; Stokes 2018).

respect to the first-person perspective—the visual information as content—and what it's like to have them. Given the richness of these similarities, it is natural to ask what distinguishes visual episodic memory from (non-veridical) visual imagery.

Following Byrne (2010: 15-16), this question divides into two. The ontological question asks, what is the difference between (visual) imagery and (visual) memory, qua mental states or kinds? The epistemic question asks, how can one determine or know whether one is (merely) image-ing or (veridically, episodically) remembering?

II. Ontology: memory, imagery, and cognitive contact

Again, visual imagery and visual memory may both be conscious (and are, in the cases that are of interest here); they both enjoy structurally and phenomenally similar content; both kinds of state are available for reflection and higher-level cognition and planning. These similarities make the epistemic question—how does one distinguish, for oneself, one's mere images from one's veridical memories—a genuinely challenging one. But first, what of the ontological question: what distinguishes these two kinds of mental state or process?

Byrne (2010) gives a plausible answer. What memory enables, while imagery does not, is *cognitive contact* with the world. Cognitive contact is just as it sounds: being in touch with some part of the world such that it can be cognized, reasoned or talked about. Byrne's claim is that perception enables cognitive contact with the present (sensible) world, and memory preserves cognitive contact with the past (sensible) world. Imagery, as such, need do neither. "Cognitive contact is the point of overlap between perception and recollection: the latter preserves the cognitive contact supplied by the former. Although imagination can involve cognitive contact—as when one visualizes the living room couch in the bedroom—it is not itself preservative: without memory, the couch would be unavailable to imagination" (Byrne 2010: 21-2).

There are a number of ways this proposal could be further characterized. John Campbell (2002), to whom Byrne partly attributes this use of the notion of cognitive contact, proposes that the substantial difference here is causal. When one moves – "non-pictorially shifts" – from mere image to a veridical memory, Campbell suggests that one experiences a kind of 'A-ha' moment, and this consists in one's recognition that one is now causally linked to some past object or event. This cognitive contact enables one to employ demonstratives to make reference to the relevant part(s) of the past world— "*that* speech really was inspiring". The difference here, importantly, is not in the mental image, but in this added self-awareness of causal link.

One might understand cognitive contact functionally. As a matter of cognitive architecture, perhaps it is the case that memories, but not mere imagery, function to preserve a connection with the past world, again, such that it can be further thought, reasoned, or talked about. This is what mental states of the memory category *do*; this is their role in the overall cognitive economy. One might even maintain that this mental kind was selected for or bestows some adaptive advantage for performing this role.

Finally, and perhaps not in a way detachable from the causal or functional characterization, one might understand cognitive contact metaphysically. Thus, it is constitutive of memories (in a way analogous to perceptual experiences) that they preserve contact with the past world. One way to put this is relationally: memories, properly understood, always involve a relation between a subject's image and an object or event in the past world. This is what it is for a mental state to be a memory. This line of thought comports well with those that take memories, as such, to be factive mental states. And all of this would be a contrast to mere images, which have no such metaphysical essence.

This provides a substantial, even if incomplete address of the ontological question. Cognitive contact, however it is further fleshed out, is a plausible mark for the distinction between imagery on the one hand, and perception and episodic memory on the other. We can grant this. However, short

of question begging, Byrne's answer to this ontological question cannot do any work in informing an answer to the second question. The second question is an epistemic one, and one in particular about self-knowledge.

This is clear in Campbell's own analysis of memory demonstratives. Campbell discusses a case where a family member is describing a window in a childhood bedroom.

> 'It was circular, with spokes running out from the centre, like the wheel of a ship', she says. As she talks, you form a vivid image of the window. The image may be correct, detailed, and reliable. Even at this stage, it seems that you could, on the strength of the image, form a demonstrative, 'that window'. Still, you cannot be said to remember the window. It may be, though, that as your sister continues talking, she finally succeeds in jogging your memory, so that you eventually say, 'Aha! Now I remember!'…After the shift, your image of the window may be exactly as before. There need be no pictorial change in the image. And it may be no more reliable than it was before. But this non-pictorial shift, whatever it is, marks the transition from your merely having an accurate, reliable, conscious image of the past window, to your consciously recollecting it. (Campbell 2002: 178-9).

The point to be gleaned is this: given the rich similarities between imagery and visual memory, it is not obvious by what method one distinguishes, for oneself, one's images from one's memories; it is not obvious what enables the "shift" that Campbell speaks of (and he intimates as much.) And this is true even if we suppose that something like cognitive contact *is* the ontological mark of distinction.[5]

III. A case study in self-knowledge: imagery versus memory

[5] It should be noted that none of the methods for self-discernment of imagery vs. memory (or self-discernment of one's own mental states, generally), as described below, are prescribed as fail-proof. There will most certainly be instances where one's attempt at the relevant self-knowledge will fail. It will be assumed, nonetheless, that making these distinctions for oneself, and thus achieving the relevant self-knowledge, is possible even if fallible. It will be further assumed that this is a genuinely interesting epistemic challenge, and for reasons concerning the phenomenal and psychological similarities between memory and imagery, *and* the most basic metaphysical difference between them (achieving vs. lacking cognitive contact with the past sensible world), all as outlined in section II. Thanks to a reviewer for pressing me on making these assumptions explicit.

How can one determine, for oneself, whether one is remembering some event, *e*, or merely visually imagining it? As Byrne puts the question, "How does one tell that one is recalling (and so not perceiving or imagining)?" (2010: 15). A phenomenological answer, traceable to Russell (1921), is that memories will involve a feeling of familiarity (which is supposed to indicate that the imaged *e* exists *at some time*) and a feeling of "pastness" (which is supposed to indicate that *e* existed *in the past*). So, with memory, there is some conscious feeling that one has seen, heard, perceived *e*, in one's past. Byrne's criticism is straightforward: the feeling of familiarity is likely to accompany cases of memory, where one is aware *that* one is remembering. But if there is a first-person question whether one is remembering or merely imaging, then the feeling would be highly defeasible as any kind of evidence. Taking a point from R.F. Holland, Byrne writes, " 'On meeting McX, he might strike me as familiar, which is a good sign that I remember him, or perhaps someone who looks very much like him. However, my well-remembered kitchen (for instance) does not likewise produce 'feelings of familiarity'… Certainly my kitchen is familiar, but that is just another way of saying that I remember it well; it is not to hint at how I know that I remember it" (Byrne 2010: 24). So however common the feeling of familiarity may be *in genuine cases of memory*, that feeling cannot itself explain how one identifies genuine memories as such. And this is at least in part because "an image might be familiar because 'you have amused yourself by creating some such fanciful image as this on many occasions in the past' (Holland 1954, p. 468)" (Byrne 2010: 23). Byrne is therefore right to urge that the feeling of familiarity may often lead one awry if a question concerns the nature of one's own mental states: whether one is enjoying a veridical episodic memory, or merely engaging non-veridical (but no less familiar) mental imagery. This critique is worth highlighting, both because it charges an elegant and intuitively plausible solution, and because Byrne's own solution fails on similar grounds.

Byrne's answer to the epistemological question employs Gareth Evans's (1982) analysis of mental self-ascription. The analysis involves an analogy between how (Byrne takes) Evans's analysis

to characterize self-knowledge of visual experience and how it may characterize self-knowledge of visual memory. Of the first, Byrne writes:

> One comes to *know that* one sees a sleeping cat by an inference from the visual world to the mind. One uses one's eyes to investigate the visual world, discovers that it contains a sleeping cat, and concludes from this premise that one sees a sleeping cat. This method will usually produce *knowledge* of the conclusion, because one would not know the premise unless one were to see a sleeping cat. Hence one knows what one sees by literally directing one's eyes "outward—upon the world" (Evans 1982, p. 225): if there is a sleeping cat there, then one may safely conclude one sees it. (Byrne 2010: 25; emphasis added)

'Visual world' is intended to denote "the world as potentially revealed by vision" (Byrne 2010: 25), thus including shapes, colours, rest/motion, depth, perhaps high-level properties like 'being a cat' but excluding, by contrast, sounds or flavours or smells. Byrne's proposed method for acquiring self-knowledge of seeing, in schematic form, therefore looks like this:

> (SKP)
> S uses *one's eyes* to investigate *the visual world*.
> If there is an x in S's *visual world*, then S sees an x.
> So, S knows that S sees an x.

Of self-knowledge of memory, Byrne writes:

> Vision reveals the present visual world, how things are visually now. …Visual recollection, in contrast, reveals the past visual world, how things were visually. One comes to *know that* one recalls a sleeping cat by an inference from the past visual world to the mind. One uses one's memory to investigate the past visual world, discovers that it contains a sleeping cat, and concludes from this premise that one recalls a sleeping cat. As before, this method will usually produce *knowledge* of the conclusion, because one would not know the premise unless the conclusion were true. (Byrne 2010: 25; emphasis added)

Byrne's analogous method for acquiring self-knowledge of visually remembering, in schematic form, therefore looks like this:

> (SKM)
> S uses *one's memory* to investigate *the past visual world*.
> If there is an x in S's *past visual world*, then S remembers an x.
> So, S knows that S remembers an x.

Now one needs to bear in mind the question of central interest here, posed to oneself: is one visually (veridically) remembering some x, or is one merely visually imaging it? Byrne's analogy is supposed to be between the two methods SKP and SKM. The analogy is elegantly simple. But it is too simple. There are at least two ways that SKM, allegedly analogous to SKP, fails. And once we see this, we will see how unhelpful this epistemological story is. (And for that matter, the degree to which SKP is successful will depend upon independent commitments about the nature of visual perception, as will be seen below).

The first problem concerns the presumed analogy between 'one's eyes' in the first step of SKP and 'one's memory' in the first step in SKM. In the vision case (SKP), it is appropriate to make one's eyes the method of inspection, since then one can at least know whether one is having a perceptual experience, in the visual modality, as of some x. (This, on a representationalist theory, is consistent with one's experience being illusory or hallucinatory. More on this complication below). But, in the memory case (SKM), the very question is whether one is imaging or remembering. And so, we can't assume that one can knowingly use one's memory to do the relevant investigation. This is nothing short of question-begging, and certainly is not going to help in any genuine case in which one presently lacks knowledge of the nature of one's mental states and wants to determine their natures. Accordingly, SKM should be adjusted to something like this:

(SKM1)
S uses *one's visualization* to investigate *the past visual world.*
If there is an x in S's *past visual world*, then S remembers an x.
So, S knows that S remembers an x.

SKM1 imports no loaded assumptions about one's knowledge of the *character* of the relevant mental states (i.e. memory or mere imagery).[6]

The second problem concerns the presumed analogy between 'visual world' in steps one and two in SKP and 'past visual world' in steps one and two in SKM/SKM1. Again, one can see why this might be appealing in the visual case: one investigates, using one's eyes, what is visibly available to one (which is different from what is auditorially or haptically available to one). And this will provide some self-knowledge: it will tell one that one is having a visual experience. Byrne wants the analogy in the memory case, since he is committed to an ontology whereby visual memory preserves cognitive contact with the world: it connects us with the past visual world, to be believed, reasoned about, and so on. But in this epistemic context, one cannot assume that, using one's visualizations, one is investigating the *actual* past visual world. This, again, would simply beg the question, assuming that one is somehow in contact with the past visible facts. So this this part of the analogy also fails. Accordingly, SKM/SKM1 should be adjusted to something like this:

> (SKM2)
> S uses *one's visualization* to investigate *the seeming past visual world.*
> If there is an x in S's *seeming past visual world*, then S remembers an x.
> So, S knows that S remembers an x.

The revision to SKM2 is forced so as to avoid question-begging and preserve the analogy with SKP. But the revised method fails (or, if one likes, the argument fails). There is little reason to think that an investigation of this kind—relying on visualizations (images) and the seeming past visual world (so, appearances as of a past reality)— is going to ensure a conclusion of self-knowledge of this kind. For all one knows, one could be inspecting, by visualizing, merely imaged objects or events.

Return, now, to the criticism of Russell's feeling of familiarity-solution. The worry there was that while a feeling of familiarity will often accompany a genuine memory, it can provide no guide to

[6] The choice of 'visualization' here is somewhat arbitrary. One might instead input 'images' or 'visual images' in the relevant slot in the first step of SKM1.

whether one is enjoying a genuine, veridical memory. As Holland put it, the mental image in question might feel familiar because "you have amused yourself by creating such fanciful image as this on many occasions in the past" (1954: 468). A similar problem undermines the viability of SKM2. An image, or an imaged object, might upon investigation through visualizing, seem to appear in one's past visual world for the simple reason that it is an image (or set of images) that one commonly or vividly entertained, perhaps for pleasure or perhaps out of some worry or anxiety. Here one might think of embellished memories: cases where one remembers that some event occurred in one's life, but lacking visual memories of some of the details one embellishes with images of one's own confabulation. Similarly, think of family stories that are told over and over again, often with subtle embellishments added with each telling (analogous to the way a gossip circle works). Or one may even think of cultural myths, folklore, or narrative fictions that prescribe rich perspectival, visual imagery. Perhaps most simply, in some cases one may have something at stake regarding past events. You and I have a disagreement, and a bet, about whether some event occurred—say whether Jose was at the party last night. My wishful thinking causes, but does nothing to justify, my conviction that my images of Jose at the party are tracking facts about the past visual world. In all of these cases, one may be unable to inspect (through) one's images and determine whether they have any contact with the past visual world and, thereby, be unable to answer the central epistemological question for oneself.

Byrne's transparency method therefore fails by itself to enable self-discernment of visually imaged non-actual events from veridical episodic memories. The method was adapted from Evans, and it is instructive to more carefully interrogate what Evans did and did not prescribe on topics of mental state self-ascription, and why the visual case is importantly different from the memory/imagery case. What is right about the transparency method is that in the case of belief or

visual experience, to determine whether one believes or visually experiences (that) *P*, one turns *outward* to the world. On this Evans writes:

> I get myself in a position to answer the question whether I believe that p by putting into operation whatever procedure for determining whether *p*. (There is no question of my possibly applying a procedure for determining beliefs *to something*, and hence no question of possibly applying the procedure to the wrong thing). If a judging subject applies this procedure, then necessarily he will gain knowledge of one of his own mental states…" (Evans 1982: 225)

Note the parenthesized sentence in this quoted passage. For all Evans says here, the transparency method could be used to investigate the wrong thing, in the following sense. In the case of visual experience, one could, as it were, investigate the hallucinated world, or the illusory world. And if one wants to know if one is veridically perceiving *or* hallucinating, one will come away with knowledge of one of one's own mental states, but maybe not all of its nature (say, veridical or not). So, the method is not prescribed by Evans as sufficient for the case of visual perception and, for similar reasons, it will fail for the imagery versus memory case.

What Evans does intend this method for, and he is right about this, is to determine the *content* or type of content of one's mental state. He is explicit about this:

> [A] subject can gain knowledge of his internal informational states in a very simple way: by re-using precisely those skills of conceptualization that he uses to make judgements about the world…He goes through exactly the same procedure as he would go through if he were trying to make a judgement about how it is at this place now, but excluding any knowledge he has *of an extraneous kind*. (That is, he seeks to determine what he would judge if he did not have such extraneous information.) The result will necessarily be closely correlated with the content of the informational state which he is in at that time (Evans 1982: 227-8).

Evans is suggesting that one turns outward to the world, and then identifies what one would judge, independent of background beliefs, knowledge, expectations. And what one gets is the content of what one would judge just on the basis of one's visual experience. One can further identify the kind of information (visual, rather than auditory or haptic). But none of this implies that one can, using this method, determine the *accuracy* and, thereby on certain theories of perception, the *character* of the

content-bearing state. Why? Because with perception, one can only determine, by this method, what the state alone gives one, say, visual information about some objects or events. This is the point of excluding "extraneous information". But use of the method does not distinguish between veridical and non-veridical visual information.

What this comes to in the visual/perceptual case depends upon one's theory of perception. Evans's transparency method, captured by SKP, suffices to enable knowledge of the informational content of one's mental state, and how that information is given to the subject, say, visually. It does not suffice, for reasons just articulated, to enable knowledge of the accuracy or veridicality of that mental state, whether one is veridically visually experiencing or merely hallucinating an *x*. On a direct realist or disjunctivist theory of perception, this entails that the method is insufficient to enable knowledge of the character or kind of mental state that one is in since, on that view, only *veridical* visual experience is genuine perceptual experience, and hallucinations are of a distinct mental kind.[7] On an intentionalist or content theory of perception, the method does suffice to enable knowledge of the character of the mental state, because illusory, hallucinatory, and veridical visual experience are all of the same kind: genuine visual perceptual experience.[8] But again, the method will fail to enable knowledge of the veridicality or non-veridicality of the content-bearing state. Therefore, how far SKP goes to address the epistemological question—self-knowledge of one's visual experiential states—varies according to independent commitments about the ontology of mind.[9]

Things are simpler, and simply worse off, for the transparency method as applied to self-knowledge of memory versus imagery. This is for the reason that in this epistemic context, the

[7] Hinton 1973; Snowdon 1990; Martin 2004.
[8] Evans himself, on some of the very same pages cited above, espouses a content view. Intentionalists include Harman 1990; Tye 1990; Dretske 2003.
[9] In this respect, and again depending upon one's theory of perception, SKP may be inadequate even to the analogous task for the visual/perceptual cases, and so not much safer than SKM for the memory/imagery cases (contrary to initial presentation above).

question is precisely whether one's mental state is veridical (episodic memory) or not veridical (mere imagery). There are no non-veridical episodic memories (even if there are, on some views, genuine non-veridical perceptual experiences). And for all the reasons just given, the transparency method (from SKM to SKM2) may reveal the content of one's imagistic state but fails to reveal whether that state represents the actual facts of one's past world.

The transparency method (understood as SKP and SKM2) enables self-knowledge of content and type of content, but not of a mental state's character or kind. Put in terms of Byrne's ontology: if one wants to know whether one remembers some event, and this amounts to preservation of cognitive contact with the past visual world, using one's visualization alone will not provide a reliable means to determine that contact. Just using one's imagery, for all one knows one is merely imagining some event as being in one's past.

IV. Towards a more complete self-knowledge: Holism and inference

The very point of the transparency method is to, as it were, "look through" the mental states directly to their content, to the events and or objects those mental states represent. This is an effective method if content or type of content is the first-person goal. However, if the or an additional goal is to identify some other fact about the mental states themselves—in particular, what philosophers have variously called their category, kind, mode, character, or attitude—then this method will fail. And this is important since distinct mental kinds can token the same content type. One can visually, veridically remember or merely visually imagine the same event, say, winning that foot race with all the neighbourhood kids on a past summer's evening. One can veridically visually perceive or suffer a visual illusion with the same content, say, as of a small body of water in the

distance. In these cases, if the inspection is really *through* a transparent mental state, then the character of the state will not appear to one on that inspection.[10]

If a theorist insists on the transparency method as the exhaustive means by which we can come to know our own minds then, given the above analysis, the consequence is scepticism. Just as one cannot (or at least not in relevant contexts), just by looking, be certain that one is not in fake barn country or that one is not hallucinating or that one is not a brain in a vat, one cannot be certain, by application of the transparency method alone, whether one is merely imagining some event or, as we say, *actually* remembering it. As with philosophical scepticism generally, once one is aware of a possibility that is incompatible with *P* (for instance, the possibility, *Q*, that one is merely imagining some event *e*, is incompatible with the proposition, *P*, that one is remembering *e*), and one cannot rule out the incompatible proposition by certain methods (in this case, the transparency method), then one's reasons or justification for believing *P* are undermined. A similar point might be put in terms of reliability. Again, if the above analysis is apt, then there are ample scenarios where application of the transparency method will fail to ensure that one correctly distinguish memory from imagery. Accordingly, the frequency with which the method produces true beliefs (say, that one is remembering *e* rather than merely imagining *e*) may be below the threshold for reliable belief-forming processes. Finally, justification and reliability to one side, it is clear that the method will sometimes produce error. One might mistakenly judge that one is remembering an event when one is merely imagining it (or vice versa). Accordingly, one lacks self-knowledge in this case. This

[10] Again, the result in the perceptual case depends upon one's metaphysics of perception. The method fails to enable first-personal discernment of perceptual experience from illusion/hallucination. For a disjunctivist, this is to say that the method fails to distinguish instances of perception from instances of a distinct mental category (or categories). For the intentionalist, the method fails to distinguish good (veridical) cases of perception from bad (illusory or hallucinatory) cases (but these are all cases of the same kind: visual perception).

suggests that the transparency method does not enable the epistemic security that is often supposed to be distinctive of self-knowledge.[11]

This sceptical result is avoided not by abandoning the transparency method, but by abandoning it as the sole or sufficient method for acquisition of self-knowledge. The general method for acquiring self-knowledge, typically contrasted with the transparency method, is some kind of introspection: one looks inward rather than outward. Descartes is one famous champion of this method, and indeed Evans cites Wittgenstein as inspiring his own transparency method as a response to Descartes: "I think Wittgenstein was trying to undermine the temptation to adopt a Cartesian position, by forcing us to look more closely at the nature of our knowledge of our own mental properties, and, in particular, by forcing us to abandon the idea that it always involves an *inward* glance at the states and doings of something to which only the person himself has access" (Evans 1982: 225). Evans then goes on to motivate his method of looking outward and, as we have already seen, this method is plausibly used when it is the content of one's own mental state (or at least contents of some kinds of states) that one wants to identify. This method falls short of enabling self-knowledge of the characters of one's individual mental states. And indeed Evans, as quoted here, allows that the transparency method might *sometimes* be supplemented with some kind of introspection or looking inwards, or that the former is not the exclusive method for self-knowledge.

Consider again our basic case, where one'e inquiry concerns a mental image of *e*, and whether this image is a veridical memory (or part of one) or is a mere image. Just as an application of the transparency method to this case fails, mere application of straightforward introspection would similarly fail. One common example of an introspective method is given by the "inner sense account" (Armstrong 1968, Goldman 2006). As its name indicates, this method involves a kind of

[11] Classic sources for this claim about epistemic security are both Descartes and Locke. Examples of contemporary theorists that make some, often modified claim, about the epistemic security of self-knowledge include Peacocke 1999; Gertler 2012, Siewert 2012. For general discussion, see Gertler 2015.

"mental scan" of one's inner states. But merely turning inwards to a scan of the image will not, by itself, provide information about the connection (or not) with the past visible world. Indeed, no combination of outcomes from the inward-looking and outward-looking method—say, the content, phenomenology, and feeling of familiarity—will reliably inform one of whether the state in question is a genuine memory or a mere image. The limitation here does not attach to inwardness or the outwardness of the inspection, but instead to what we might call the *isolationism* of either method, when applied in this way. What's needed to make the right self-identification is something more holistic.

Peter Carruthers' recent interpretive sensory-access account (ISA) takes scepticism and the risk for error (indeed, genuine patterns of error) seriously (Carruthers 2011). Carruthers' view, most simply, is that we have a single mechanism for knowledge of minds: what most readers will recognize as our capacity for *folk psychological mindreading*. We might determine contents of sensory states via straight application of transparency (looking outward), and might also identify the phenomenology of affective states via introspection (looking inward). However, Carruthers argues that identification of the wider array of mental kinds—including propositional attitudes, and in some cases their contents—takes application, to oneself, of the same interpretive methods invoked for explaining and predicting the minds and behaviours of others. When employing this method for others, one takes in all relevant sensory input or data, draws upon background information and learning, identifies relevant contextual factors, and applies mental concepts. Thus, the method is both interpretive and sensory.

This is not the place for full exegesis or analysis of Carruthers' ambitious theory. It is the place to glean certain important contrasts between this account and the others considered. These features may not be ones that Carruthers himself espouses or highlights, but they provide some general schematic lessons for an analysis of a more complete self-knowledge. Again, it is central to

the ISA account that one employs, for self-knowledge, the same method used for knowledge of others' minds. Setting to one side orthogonal debates about folk psychological mechanisms and accuracy, this much should be uncontroversial: determining the mental states (and thereby explaining the behaviour) of others is not done, as it were, in informational isolation. One rarely determines what mental states to attribute to the target agent in piecemeal or one-off fashion, that is, determining first that *S* believes *P*, and *then* that *S* desires *Q*, and *then* that *S* fears *R*, and so on. Instead, these determinations are typically made in connection with one another and in ways sensitive to the coherence of the candidate mental states (those states to-be attributed). These determinations are also made in ways sensitive to environmental circumstances (what can *S* see, hear, touch…?), to background knowledge (how do people typically think or behave in a context like this?), and to mental concepts (what kinds of mental states dispose one to act in such-and-such a way?). Although a term with a loaded philosophical history, this method is *holistic*, and dramatically so by comparison to the isolationist methods described above.

The method is also inferential, at least in the following way. Once one has selected the various bits of sensory, environmental, and conceptual data, an inference is drawn about the agent targeted for explanation. This may not be an inference in syllogistic or like form, but it is a conclusion based on and supported by a variety of evidence. Put simply, and as a point not about phenomenology but about the causal structure of the mindreading: one does not *just see* that the subject is doing such and such because she has such and such mental states. One might, given a sufficiently familiar situation, describe one's mindreading in these terms of phenomenal immediacy, but the structure of the underlying process is complexly mediated.

Now to apply—and this is at most a sketch—these lessons to the case of self-knowledge. What we can learn from the ISA account is that broad self-knowledge must be holistic *and* inferential (at least in the weak sense just discussed). Again, taking the memory versus imagery case

as a test, a method like the ISA prescribes first that one does not inspect (inwardly or outwardly) a mental state or content in isolation. One attempts to determine how that state or content relates to others (where other states will be gotten at by looking outward to contents represented, or by looking inward to other representations and their phenomenologies). Does or how does the target state cohere with other mental states of one's own? Is it consistent with other states drawn up in working memory? Are there identifiable causal connections between the states/contents considered? One further applies general knowledge of minds, about mental kinds and how they relate to other kinds and to behaviour, about relations between environmental situations and mental representation, and so on.

Suppose one is considering an image of an event *e*: Jose at last night's party. This is no philosopher's example: surely one may have a vivid visual image of Jose at the party but be uncertain whether one is remembering or merely imagining this event. Both the transparency method and internal sense method, in isolation, fail for the above reasons. The holistic method in this case will make appeal to a variety of additional information. How does this image cohere with other images that one can easily draw to mind (perhaps one has images of Jose in only one of the contexts of the party, while having many vivid images of all of those same contexts)? How does it cohere with relevant background knowledge (perhaps one knows that Jose is terribly socially anxious and avoids parties at all costs)? Is this image consistent with other occurrent mental states (perhaps one believes that Jose was at a conference in London this week)? Here again the procedure is inferential (or if one prefers, quasi-inferential). The outcome of this evidence-gathering procedure, however rapidly it might be applied, is a conclusion about the targeted mental state, in this case, one's image of *e*. One can hereby infer that one is, let's suppose, merely imaging *e*. When performed well, the holistic method warrants the belief inferred. And when the consequent belief is true, the method enables self-knowledge: one can know that one is just imagining *e*.

If roughly correct, this sketched approach suggests that some common, traditional assumptions about self-knowledge are mistaken. Knowing one's own mind is not done infallibly: we make errors, even about our own mental states and processes. And by the same token, we do not have "omniscient" access to the nature or contents of our mental states. Whether this access is more epistemically secure than access to facts about the world outside our minds is an open question. But one thing that is clear is that self-knowledge is not, as ostensibly assumed by Hume and many others, perfectly epistemically secure. These lessons encourage scepticism only if we insist on a singular method of acquiring self-knowledge, and moreover one that is applied in isolation. Instead, the sketch here suggests that a pluralism about method, applied holistically to the mind, is the strategy for avoiding error and, in the good cases, knowing one's mind. Finally, these lessons were learned, following the important work of Byrne and Carruthers, respectively, by moving beyond sole emphasis on the doxastic state of belief and abandoning the assumption that self-knowledge must achieve causal or structural immediacy. Knowing more of one's mind requires using more of one's mind.

Acknowledgements:

Early versions of this paper were given at the Imagination and Fiction Workshop at the University of Konstanz in 2016, and the 2019 Annual Meeting of the Southern Society for Philosophy and Psychology. Thank you to the audiences at these events, especially to Magdalena Balcerak Jackson, Amy Kind, Julia Langkau, Peter Langland-Hassan, Shen-yi Liao, Kathleen Stock, and Neil Van Leeuwen. Thanks also to the editor of this volume and a helpful reviewer.

Bibliography

Armstrong, D. (1968/1993). *A Materialist Theory of the Mind*. London: Routledge.

Byrne, A. (2010). Recollection, Perception, Imagination. *Philosophical Studies* 148: 15-26

_____.(2011). Self-knowledge and transparency. *Proceedings of the Aristotelian Society* 85(10: 201-21.

Campbell, J. (2002). *Reference and Consciousness*. Oxford: Oxford University Press.

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Cassam, Q. (2007). Ways of Knowing. *Proceedings of the Aristotelian Society* 107: 339-58.

Dretske, F. (2003). Experience as Representation. *Philosophical Issues* 13: 67-82.

Evans, G. (1982). *The Varieties of Reference*. Oxford: Oxford University Press.

Gertler, B. (2012). Renewed Acquaintance. In D. Smithies and D. Stoljar (Eds). *Introspection and Consciousness*. 89–123.
_____.Self-Knowledge. *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition). E. N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2015/entries/self-knowledge/>.

Goldman, A. (2006). *Simulating Minds*. Oxford: Oxford University Press.

Harman, G. (1990). The Intrinsic Quality of Experience. *Philosophical Perspectives* 4: 31-52.

Hinton, J. M. (1973). *Experiences*. Oxford: Clarendon Press.

Holland, R. F. (1954). The empiricist theory of memory. *Mind* 63:464–486.

Hume, D. (1739–1740/1978) A Treatise of Human Nature, L. A. Selby-Bigge (ed.); revised by P.H. Nidditch, Oxford: Oxford University Press.

Martin, M.G.F. (2004). The Limits of Self-Awareness. *Philosophical Studies* 120: 37–89.

Nanay, B. (2015). Perceptual content and the content of mental imagery. *Philosophical Studies* 172(7): 1723-36.

Peacocke, C., 1999, Being Known, Oxford: Oxford University Press.

Siewert, C. (2012). On the Phenomenology of Introspection. In D. Smithies and D. Stoljar (Eds). *Introspection and Consciousness*. 129–168.

Snowdon, P. (1990). The Objects of Perceptual Experience I.' *Proceedings of the Aristotelian Society*, Supplementary Volume LXIV: 121–150.

Stokes, D. (2018). Mental Imagery and Fiction. *Canadian Journal of Philosophy* 49 (6): 731-754. 2018

Stokes, D. and Biggs, S. (2015). The dominance of the visual. In D. Stokes, M. Matthen, and S. Biggs (Eds). *Perception and its Modalities*. New York: Oxford University Press. 350-78.

Russell, B. (1921/1995). *The analysis of mind*. London: Routledge.

Tye, M. (1990) *Ten Problems of Consciousness*. Cambridge, MA: MIT Press.

Williamson, T. (2000). *Knowledge and its Limits*. Oxford: Oxford University Press.