Self-Knowledge and Imagination

Peter Langland-Hassan[*]

Department of Philosophy, University of Cincinnati, Cincinnati, OH, USA

Abstract: How do we know when we have imagined something? How do we distinguish our imaginings from other kinds of mental states we might have? These questions present serious, if often overlooked, challenges for theories of introspection and self-knowledge. This paper looks specifically at the difficulties imagination creates for Neo-Expressivist (Bar-On 2004), outward looking (Byrne 2005, 2011, 2012b), and inner sense (Goldman 2006, Nichols and Stich 2003) theories of self-knowledge. A path forward is then charted, by considering the connection between the kinds of situations in which we can reliably say that another person is imagining, and those in which we can say the same about ourselves. This view is a variation on the outward-looking approach, and preserves much of the spirit of Neo-Expressivism.

Keywords: imagination; self-knowledge; introspection; neo-expressivism

1. Introduction

I am imagining that there are zombies. How do I know that I am? How do I know I am not merely supposing that there are zombies, or wishing that there are, or judging that there are? Of all the attitudes I could take toward the proposition that there are zombies, how do I know it is the attitude of imagination? And how do I know it is the proposition that there are zombies, and not some other that I am imagining?

Now I am imagining a zombie (an object, not a proposition). How do I know that I am? Might I be remembering one? Or seeing one? Or imagining something else entirely?

Part of the reason I am interested in these questions is that they are not as pressing for most other kinds of mental states. Some promising accounts have been given of the special sort of knowledge we have of our own beliefs and desires, for instance. Here I have in mind the Neo-Expressivist (Bar-On, 2004) and outward-looking (Byrne 2005, 2011) approaches, discussed

---

[*] Email: Langland-Hassan@uc.edu

below. But imagination seems to pose special problems. In what follows I discuss the challenge imagination presents to the Neo-Expressivist, outward-looking, and inner sense/internal monitoring (Nichols and Stich 2003, Goldman 2006) views of introspection and self-knowledge.

After assessing these three views (in Sections Two through Four), I offer some proposals for how to move forward, in Section Five. An explanation of how we can attain knowledge of *many* our own imaginings is available, I suggest, if we accept a broad notion of imagination— one that counts as imagination all elaborated cognition that, in a sense to be explained, represents situations beyond what we believe to be the case. Here I build on Byrne's outward-looking approach to self-knowledge, while offering a view that preserves much of the spirit of Neo-Expressivism. A remaining difficulty, however, is to explain our knowledge of imaginings that perfectly align with our existing beliefs. With respect to these, the most promising path may be to pursue a redescription of such cognition in mental state terms for which the question of self-knowledge is more tractable.

1.1 Self-knowledge in a quiet room

Questions of self-knowledge are questions about our knowledge of our own minds. I will limit my discussion here to our knowledge of our own current mental states (such as our beliefs, desires, and intentions), setting aside the question of how we know our own personality traits (such as cowardice or generosity). Discussions of self-knowledge typically focus on explaining two features of our knowledge of our own mental states. The first is the special epistemic *security* this knowledge seems to have. Many have thought that our beliefs about our own current mental states are far less likely to be in error than our beliefs about other contingent matters. If that is true, we would like some explanation of why such beliefs are so secure.

The second aspect of self-knowledge that calls for explanation is the special sort of *access* people seem to have to their own mental states.   We each seem able to know our own current mental states in situations where others cannot. This is true whether or not we view the increased knowledge as being especially epistemically secure, relative to our knowledge of other matters. The question of access is thus a separate question from that of epistemic security, even if there may be important connections between them (Cf. Byrne's (2005) distinction between privileged and peculiar access).

It is the access question that will be my focus here. We each seem able to know what we are imagining in many cases where an outside observer could not. I will use the notion of a *quiet room case* to highlight this discrepancy. Jane, sitting motionless in a quiet, empty room, is imagining that it is snowing. And she is in a good position to know that she is. At least, so much is commonly assumed. A friend peering in at her through a window, or sitting next to her, is not in a good position to know what she is imagining. Jane's is thus a *quiet room* case of self-knowledge. Explaining quiet room cases where a person knows she is imagining (and knows what she is imagining) is a particular challenge for theories of self-knowledge.

1.2 *Propositional imagination and sensory imagination*

As a last opening remark, it is common to distinguish between (at least) two forms of imagination: propositional imagination, and sensory (or "perceptual") imagination. Propositional imagination occurs when a person imagines *that p*. Like other propositional attitudes (e.g., beliefs and desires), propositional imagining is often assumed not to require the use of any sensory imagery, even if it is sometimes accompanied by such imagery. Propositional imagination is typically thought to be the cognitive component of pretend behavior: a person pretends that *p* partly by means of imagining that *p* (Nichols and Stich 2000, Carruthers 2006, Currie and Ravenscroft 2002).[2] If engaging in pretense does not require one to generate sensory imagery, then neither (one might suppose) does propositional imagination.

Sensory imagination, on the other hand, is typically understood as requiring (as a necessary condition) endogenously-triggered mental imagery, keyed to some sense modality (Gendler 2005, Noordhof 2002, Peacocke 1985, Byrne 2007). An important question with respect sensory imagination, so defined, is whether all cognition involving mental imagery is to be considered sensory imagination (including, e.g., episodic memory), or whether sensory imagination is best understood as a subset of all mental imagery-involving cognition. I will come back to this question about sensory imagination later (in Section Three).

---

[2] Some will go further by characterizing propositional imagination as involving the taking of a distinct cognitive attitude toward a proposition, where this draws on elements of cognitive architecture over and above those governing other attitudes like belief and desire. However, I argue elsewhere (Langland-Hassan, 2012) that propositional imagining is a way of using one's standing beliefs to arrive at new judgments, and that doing so need not involve elements of cognitive architecture over and above one's ordinary beliefs and desires. Hence my more neutral characterization of propositional imagination here.

In this paper I will be discussing both propositional and sensory imagination. For the most part, I will use 'imagination' broadly to refer both to propositional and sensory imagination, including within sensory imaginings all mental imagery-involving cognition. However, I will occasionally specify that my remarks apply only to either propositional or sensory imagination.

2. Questions for Neo-Expressivism

Dorit Bar-On's (2004) discussion of self-knowledge centers on the observation that certain claims people make about their own current mental states (what she calls "avowals") have a special authority. In most cases, questioning whether they are true, or querying their grounds, seems inappropriate. The special authority of avowals, Bar-On contends, is to be explained by their having two different expressive roles simultaneously. "A self-ascription such as 'I am scared of this dog' can be semantically about the subject and a state of her," explains Bar-On, "but can nevertheless also be seen as giving voice to the speaker's fear." Insofar as the self-ascription also "gives voice" to, and thereby expresses, the speaker's condition, "it should indeed seem inappropriate to ask after the reasons she has" for it (260-263). This focus on the expressive nature of avowals has led Bar-On's view to be known as *Neo-Expressivism*.

Bar-On's idea is that an avowal's epistemic security is guaranteed by the fact that it is simply expressing the state it self-ascribes (and to which it thereby "gives voice") (264). For instance, when I exclaim, "I am so mad at him!" I ascribe to myself a state of anger. But this self-ascription may have been uttered not as a means to reporting my state of mind, but as a means to *expressing* that anger. Why think that the statement expresses the anger, as opposed to simply reporting it? The reason is that there are *other* statements that quite obviously express the anger—such as "He's such a jerk!"—that could have been used to the same effect, in this context. (When two statements can be used to the same effect, in a context, we can say they are *performance-equivalent*). Bar-On's point is that "He's such a jerk!" and "I am so mad at him!" are performance-equivalent in this context just because they express the same mental state (my anger at him). This is true even if there is *another* sense in which the two statements express quite different propositions. In what Bar-On calls the *semantic sense,* the statement "He's such a jerk" ascribes jerk-hood to an individual, while "I am so mad at him!" ascribes anger to myself. It is in what Bar-On terms the *action sense* that both exclamations expresses my anger (2004, 219-220). Avowals, on Bar-On's account, are thus expressive acts where we express, in the

action sense, mental states that are self-ascribed in virtue of what is semantically expressed by the same communicative vehicle. Thus, "I am so mad at him!" is an avowal just because it expresses, in the action sense, the anger that it self-ascribes, in the semantic sense.

An important part of this picture is that avowals are performance-equivalent (or nearly so) to certain non-introspective statements that simply express (and do not self-ascribe) the state in question. This is what warrants the claim that they in fact express, in the action sense, the states they also self-ascribe. Other examples of this sort of performance equivalency include the avowal, "I believe that *p*," which can be seen as (nearly) performance-equivalent to the simple declaration that *p*. For there are many contexts where one might say either "*p*" or "I believe that *p*" simply as a means to expressing the belief that *p*. And a self-ascriptive avowal such as "I want the Yankees to *lose*!" may be performance-equivalent, in a context, to an exclamation such as "Boo Yankees!" which simply expresses the desire ascribed by the avowal. The self-ascriptions made through an avowal can then be seen as especially secure just because they are directly caused by, and therefore expressive of, the ascribed state itself. There was no need for an inner monitoring mechanism to accurately detect the ascribed state prior to the self-ascription.

In addition to offering leverage on the question of the special authority of avowals, the Neo-Expressivist provides an elegant means for tackling the question of introspective access. Instead of appealing to a special sense-like faculty of inner observation, the Neo-Expressivist can hold that it is simply by appreciating the kinds of performance-equivalencies just mentioned that a person learns to make true avowals. For instance, in order to accurately ascribe a desire to oneself, one simply needs to grasp that situations where one would say (for instance) "Go Red Sox!" are situations where one can correctly say "I want the Red Sox to win!" to the same effect. In general, because avowals express (in the action sense) the states they self-ascribe (or semantically express), can one learn to reliably make self-reports concerning those mental states simply by latching on to the relevant performance equivalencies between such reports and certain first order, non-self ascriptive statements.[3] In this way, the Neo-Expressivist is able to explain many cases of self-knowledge by appeal only to a language-learning capacity we in any case know ourselves to possess. Note also that the Neo-Expressivist is not committed to holding

---

[3] As a referee notes, questions can be raised about whether this sort of pragmatic understanding of performance-equivalence is sufficient for understanding that one is in the mental state one self-ascribes through an avowal. For present purposes, however, I will grant the Neo-Expressivist view of self-knowledge with respect to these cases, in order to show how imagination presents difficulties nonetheless.

that we only can make the relevant avowals when speaking *aloud*. As Bar-On and Chrisman (2009, 138) point out, avowals can occur silently, in inner speech, as when one says to oneself "I wish it were not so cold!" as a means to inwardly expressing one's wish for warmer weather. Here the inner speech utterance both expresses the wish, in the action sense, and reports it, in the semantic sense. In the process of inwardly expressing the wish, in the action sense, one expresses a judgment, in the semantic sense, that qualifies as an instance of self-knowledge.

The problem with imagination, however, is that avowals of imaginings—statements such as "I am imagining that p," or "I am imagining an *x*"—are not obviously pragmatically equivalent to anything that would count as a (non-introspective, non-self-ascriptive) *expression* of an imagining. And if there is no way to self-ascribe an imagining by making a statement that also expresses the imagining, the Neo-Expressivist account of how we learn to securely self-ascribe mental states cannot be extended to imagination. The question of access is left unexplained.

To see this, we can first ask what might count as a non-self-ascriptive expression of an imagining. I will focus on propositional imagining here, since propositions are more naturally related to episodes of linguistic expression. If simply asserting "p" expresses, without self-ascribing, the belief that *p,* what sort of statement would express, without-self-ascribing, an imagining that p? It is not immediately obvious how one might occur. On reflection, pretense seems to me to offer the most plausible context where one might occur. Suppose two children are pretending to be museum guards. One says to the other: "The jewels are missing!" Perhaps this statement is an expression of the child's state of propositionally imagining that the jewels are missing. (It is not an expression of a belief, after all). Notice, however, that here there is no expressive equivalence to the self-ascriptive avowal, "I am imagining that the jewels are missing!" That is, it is not plausible that "I am imagining that the jewels are missing!" serves to express, in the action sense, the imagined proposition that the jewels are missing. There is not a parallel to the way in which "I hope the Yankees *lose*!" plausibly expresses (in the action sense) the hope that the Yankees will lose. For the child to say, during the pretense, that she is imagining that the jewels are missing would be for her to step outside of the role she is playing in the pretense. She would then be speaking for *herself*—saying that she herself is imagining—and not for and about the character she is playing in the pretense, who presumably is *not* imagining. Moreover, even if she were interpreted as speaking for the pretend character in saying "I am

imagining that the jewels are missing!" that statement would not be equivalent to the character's simply saying "the jewels are missing!" For it would imply that the pretend character is only imagining that the jewels are missing, not that she really believes them to be missing. Thus, "I am imagining that the jewels are missing!" is not pragmatically equivalent, in the context, to saying, "The jewels are missing!" So, even if pretense is an ideal scenario where an imagining might be linguistically expressed, an avowal of such an imagining does not here "play the same role as is played by more direct expressions of one's own intentional states" in the way that Bar-On thinks guarantees the authority of avowals for intentional states (2004, 219).

In response, a Neo-Expressivist might reply that we simply have not identified the right sort of context where the performance-equivalence holds. A solitary pretense might seem to offer a more plausible context. Suppose a child is pretending to be a space alien while her parents look on. Waddling around in the way aliens do, she might say either "I'm an alien!" or "I'm imagining that I'm an alien!" Both utterances appear pragmatically equivalent, in the context. However, the seeming equivalence is shallow and subject to interesting limits. Suppose the child had altered her voice so as to sound more like an alien (just as children typically put on a "robot voice" when pretending to be a robot, or a deep voice when pretending to be a bear). When speaking in the alien voice, it would *not* be pragmatically equivalent for the child to say either "I am an alien" or "I'm imagining that I am an alien." The self-ascriptive form, said in the alien voice, conveys that, as part of the pretense, the *alien* is imagining that it is an alien, and not simply that the child is imagining that she is an alien. This suggests that, to the extent the performance equivalence holds, it is only in the context of something less than a fully engaged pretense—one where we may in any event question whether she is really imagining.

So, imagination does not sit comfortably within the general Neo-Expressivist approach to explaining both the authority of avowals, and the means by which one learns to reliably make avowals. I only say that it does not "sit comfortably" with Neo-Expressivism for two reasons. First, the door remains open to the Neo-Expressivist to propose some other context where the right kind of performance-equivalence holds with respect to imagination. And, second, I will later propose a way of accommodating imagination to a view that, if not exactly Neo-Expressivism, retains much of its spirit.

For the time being, the point remains that Neo-Expressivism identifies an interesting feature of *much* self-knowledge, in noting the performance-equivalence of certain statements that

express a mental state with others that self-ascribe one. This important idea reappears, in a different guise, in the outward-looking accounts discussed below.

3. Questions for Outward-Looking Accounts

In a now familiar passage, Evans (1982, 225) observes that, in order to answer the question of whether one believes that *p,* one need not turn one's attention to the contents of one's own mind (one need not "look inward"). One's eyes (and attention) can, so to speak, be "directed outward" in considering whether *p* is the case. Alex Byrne works forward from this idea in offering the epistemic rule BEL as procedure for generating knowledge of one's own beliefs. BEL is a rule of inference that says: if *p,* believe that you believe that *p* (Byrne 2005, 93-95)*.* BEL is a *neutral* rule, in Byrne's sense, insofar as following it does not require prior awareness of one's own mental states. To follow this rule (and so to know whether you believe that *p*), you simply need an ability discern whether *p*, and a disposition to infer from *p* that you believe that *p.*

As with Neo-Expressivism, a virtue of the outward looking approach is that it promises to explain self-knowledge, and its presumed epistemic security, without positing a *sui generis* faculty of introspection that would itself stand in need of explanation. It makes do with faculties and capacities of which we have an independent understanding and reason to believe in outside of questions about self-knowledge—*viz.*, an ability to make judgments about the external world, and the ability to learn and follow certain "neutral" rules of inference. Byrne has extended this approach of invoking neutral, outward-looking epistemic rules to explain self-knowledge of other attitudes and mental states, including desire (2012b), seeing (2012a), and thinking (2011).

In the case of desire, he offers DES.

DES: If φ-ing is a desirable option, believe that you want to φ (Byrne, 2012b). Here the idea is that judging an option to be desirable does not require a prior awareness of one's own mind. It is true that an option's being desirable is a subjective, relational matter. Yet so is the attractiveness of a face, or the warmth of a room. If judging a face to be attractive, or a room to be hot, does not require a special introspective capacity, then neither, arguably, does judging an option to be desirable. And it seems that, in most cases where an option is judged desirable,

the person doing the judging could correctly infer that she wants the option.[4] If so, we have an outward-looking route to self-knowledge for desire, on a par with BEL.

Some will object that such patterns of inference are illegitimate. After all, why should the fact that there is a squirrel on the roof make it reasonable for me to infer that I believe that there is a squirrel on the roof? Are there not many squirrels, on many roofs, of which I have no inkling? Nevertheless, as a procedure for safely generating true beliefs about my own beliefs, BEL works[5]: if, in cases where I judge that p, I move from p to the belief that I believe that p, I will not have fallen into error—even if I could not safely infer from p that *anyone else* believes that p. As Byrne emphasizes, the unusual nature of the inference—that it can only safely be made with respect to the beliefs of the person following the rule—is part of what explains the special sort of access each person has to her own mental states.

Interestingly, Byrne's BEL rule in many ways echoes the nice developmental story of Neo-Expressivism. According to the Neo-Expressivist, one begins to apply mental state terms to oneself by noting, for example, that usually when one says "p" one could, without rebuke, alternatively say "I believe that p." This discovery is in effect internalized in the form of BEL: one learns that situations where p are situations where one can, without rebuke, judge that one believes that p.[6] Whereas Neo-Expressivism emphasizes an equivalence of performance conditions between certain introspective and non-introspective statements, the outward-looking approach highlights an equivalence of inference conditions, such that a proposition pertaining to one's own mind can be rightly inferred by a person whenever that person can rightly infer a certain proposition about the external world.

The problem in the case of imagination, however, is that there do not seem to be any states of the external world from which we can safely infer that we are imagining that p.

---

[4] Byrne (2012b) qualifies and amends DES to account for several objections (such as that we sometimes do not want something we judge to be desirable). Here I aim only to give the flavor of his approach, not defend it in detail.

[5] Some will object the kind of reliability guaranteed by BEL is too brute to count as *knowledge*. Byrne considers this objection in his own work. I admit to having a fairly austere conception of what knowledge requires. But my interest, in the end, is in how we *in fact* arrive at the kinds of beliefs we have about our own mental states—beliefs others are rarely apt to challenge—whether or not the processes by which we do so satisfy intuitions about what constitutes knowledge.

[6] Learning to proceed in this way is not the same as coming to believe the conditional: if p, then I believe that p. To believe that conditional is close to believing in one's omniscience! For it is one thing to follow a procedure of the form, "If p, believe that you believe that p." It is another to have the belief: "If p, then I believe that p." The first is not a belief, but a rule one can either follow or not. Compare: "If you go swimming, wear your bathing suit" is a rule one might follow with respect to when to wear a bathing suit. But one could follow that rule without having the belief: "If I am swimming, then I am wearing my bathing suit."

Directing our attention outward, we gather few clues about what we are imagining. Or so it seems.

But consider pretense again. Young children (between the ages of two and three) are able to recognize pretense in others, and themselves, before they pass standard false belief tasks[7] (Harris and Kavanaugh 1993, Rakoczy and Tomasello 2006). The reason for this discrepancy is that recognizing pretense does not *require* an understanding of mentality. Pretense can potentially be recognized by the young child as a kind of game, where various cues (nods, smiles, special tones of voice) indicate that it is permissible to act as though something is the case that is not the case (Rakoczy, Tomasello, and Striano 2004, Lillard and Witherington 2004, Sobel 2009, Lillard 1994, Langland-Hassan 2012). Now suppose that a person can, in fact, reliably judge herself to be pretending that *p* simply by attending to these outwardly available cues. She could then follow a neutral, outward-looking rule for coming to know her propositional imaginings. The rule could take the form: If I am pretending that *p,* believe that I am imagining that *p.* After all, in cases where we one is pretending that *p*, utterances of "I am imagining that *p*" are not met with rebuke. So it seems safe to infer that one is imagining that *p* whenever one finds oneself to be pretending that *p.*

However, even granting this proposal, we would still lack a plausible outward-looking approach to explaining quiet room cases—cases where we engage in imagination without using it to guide a pretense, or any other outward action. To address quiet room cases from an outward-looking point of view, we might consult Byrne's (2011), where he aims to explain, in terms consistent with the outward-looking approach, how we know what we are thinking. Byrne uses 'thinking' here not to refer to cognition in general, but "in roughly the sense of 'a penny for your thoughts'," picking out mental activities "like pondering, ruminating, wondering, musing, and daydreaming" (2011, 105). As is the case with imagining, we are very often in a situation where there are no obvious external markers from which we can reliably judge the particular contents of such thoughts, or even whether such cognition is occurring.

Byrne's approach here hinges on our ability to discriminate instances of inner speech and visual imagery from their related acts of genuine perception. Supposing that both inner speech

---

[7] The issue of when infants and children understand false belief remains controversial. Some argue that the differential looking times of infants in studies modelled on the traditional false-belief task are evidence that an implicit awareness of false belief arises in infants as young as 15 months old (Onishi and Baillargeon 2005).

episodes and visual imaginings represent external-world properties—such as the sounds of words or sentences, in the case of inner speech, and colors and shapes, in the case visual imagery—the fact that they do so in a representationally "degraded" (i.e. less fine-grained) way offers a potential means for distinguishing each from its corresponding perceptual act. In effect, Byrne proposes that a discrimination between inner speech and genuine speech perception can be made while still "looking outward." (He makes the same point with respect to visual imagery and visual perception.) If that is right, then we can infer from objects and utterances represented in a degraded imagistic way, characteristic of inner speech and visual imagery, that we are *thinking about* those objects or utterances—even if we may not know the exact *attitude* with which we are thinking about those things (more on this below).

Determining that an inner speech utterance has occurred is, in Byrne's terms, determining that "the inner voice speaks." He proposes the rule THINK as a means to knowing when one is thinking about *x*:

THINK: If the inner voice speaks about *x,* believe that you are thinking about *x* (2011, 117).

Following THINK, one can, in essence, infer from a much-degraded "hearing" (in inner speech) of the phrase, "the Earth has one moon," that one has just thought about the Earth and the moon (Byrne, 2011, 116-121). On this approach, episodes of inner speech and visual imagery—conceived of *as kinds of mental states*—are not objects of our immediate awareness. Rather, they are mental states we enter into, in virtue of which we become aware of certain (often uninstantiated) external properties, such as the sound of a verbal utterance (in the case of inner speech) or colors, shapes, and relative spatial relations (in the case of visual imagery). THINK is thus offered as a "neutral" rule, of the kind discussed above. In this way, the view aims to avoid explaining knowledge of one kind of mental process (thinking) in terms of our prior knowledge of another kind of mental process (inner speech, or visual imagery), and to avoid a possible regress.

One could, however, reasonably question whether THINK is indeed a neutral rule. It may seem that an ability to determine that the inner voice is speaking already presupposes introspective awareness of one's own mind. Here is a way of sharpening the worry: it is possible to have a degraded auditory-phonological representation of speech while taking it to represent another's actual utterance, as in dimly hearing a whisper, or a radio left on at low

volume. Hearing an utterance *as* an utterance by the inner voice involves additional assessments, such as that the utterance does *not* derive from another's agency, that it is inaudible to others, and that one's hearing of it will not change for better or worse as one moves around. Arguably, in judging a represented utterance to have these features, one is, in effect, already judging it to be a *mental* event. For what sort of non-mental event is both an utterance *and* has these features? If judging the inner voice to have spoken is *ipso facto* judging oneself to be in a certain mental state, then following THINK presupposes a prior awareness of one's own mental states after all.

The outward-looking theorist thus needs to explain how a represented speech utterance can be determined to have these characteristically mental features (e.g., being inaudible to others, not changing in law-like ways with one's movements, not being the result of another's agency) without one's needing a special faculty for internal observation. To that end, the outward-looking theorist can propose that there are pre-conscious processes by which an auditory phonological representation is determined either to lack or possess the features that distinguish inner speech episodes from cases of degraded speech perception. It could be at this point in processing that the disambiguation between whether the utterance was inner- or outer-caused (and audible or inaudible to others) could be made. So long as those disambiguating processes did not themselves presuppose awareness of one's own mental states, one might arrive at a judgment that the inner voice is speaking without having made a prior judgment that one was in some other mental state. The regress would have been stopped.

In considering this possibility, I have in mind the kind of comparator and prediction mechanisms posited within motor-control theory, whereby predicted sensory input is compared to actual sensory input to determine whether there is a match (Miall et al. 1993, Wolpert, Ghahramani, and Jordan 1995, Hohwy 2013). This sort of architecture is thought to enable the swift correction of bodily movements, and to allow for a pre-conscious determination of whether changes in sensory input were caused by one's own movements, or the movement of something in the environment. These processes are not "introspective" in the relevant sense, lest raising one's arm be thought to require self-knowledge or introspection.[8] For they are held to govern and facilitate control of ordinary bodily movement both in humans and simple organisms, such as flies and fish (Sperry 1950, von Holst and Mittelstadt 1950/1973). I have argued elsewhere

---

[8] Though see Shea (2014) for an argument that such processes are "introspective" at least in the sense that they are metarepresentational.

(Langland-Hassan 2008) that, with some amendments, there is reason to think the same sort of architecture underlies our ability to distinguish our own inner speech from cases of hearing speech. If that is right, then there is a way of explaining how an utterance represented by an inner speech episode can represented as having paradigmatically mental features, without use of a faculty for inner observation. Thus, even if THINK is not quite neutral as it stands, following it may not require anything over and above the kind of processes at work in ordinary perception and rational inference. This seems to give the outward-looking theorist everything he is looking for.

Nevertheless, even granting Byrne's view when it comes to thinking in general, there remain difficult questions with respect to imagination. As noted above, 'thinking,' in Byrne's use of the term, is meant to capture many forms of occurrent thought, including "pondering, ruminating, wondering, musing, and daydreaming." Propositional imagining is (arguably) just one form of such thought, to be distinguished also from judging, wishing, remembering, deciding, supposing, and so on. Any of these could be what we are up to when we detect inner speech or visual imagery that concerns a particular proposition or scenario. And Byrne's account of how we know that we are thinking offers few hints as to how we might know that we are *imagining*—as opposed to judging, wondering, remembering, entertaining, intending, and so on—the proposition in question.

One way to respond, while hewing to Byrne's method, is to consider whether there is some fact, in addition to the inner voice's uttering a phrase, the detection of which could serve as a tell-tale sign that one is imagining the proposition expressed (or represented) by an inner speech episode. For instance, Byrne offers a rule called 'THINK THAT', which aims to show how we can know we are thinking that p, as opposed to simply knowing that we are thinking about x. Thinking that p, in Byrne's sense, involves assenting to the truth of p (as in: "I think that rain is likely."). Here he melds THINK and BEL:

> THINK THAT: If the inner voice says that p and p, believe that you are thinking that p
> (2011, 121).

'And p' is added to THINK to ensure that the person following the rule in fact believes that p. This way the follower of THINK THAT who hears the inner voice say that p can reliably judge that she is thinking *that p* and not, for instance, merely wishing that p.

The difficulty in the case of propositional imagination, however, is that the facts about whether we are imagining that *p* do not seem well-tracked by any such judgments about the external world. It seems we can imagine that *p* while believing that *p,* while disbelieving that *p,* or while having no opinion either way about the truth of *p.* And we can imagine that *p* while desiring that *p*, or fearing that *p,* having mixed emotions about *p,* and so on. In short, unlike most other attitudes, the facts about whether we are imagining a proposition do not seem constrained by what we make of the proposition itself. There are few if any facts about our relation to the world that we can rely upon as a means to bootstrapping our way into knowledge of our propositional imaginings. Without narrowing the broader circumstances in which one can (or cannot) imagine a proposition, it is very hard to see how a reliable outward-looking rule can be formulated.

What about the case of sensory imagination? Is the outward-looking approach better suited to explaining how we know when we are sensorily imagining something? Here it depends on how we understand the scope of sensory imagination. If any mental process at all involving mental imagery counts as an act of sensory imagination, then the outward looking approach holds some promise. It is possible that a person might come to know she is having some visual imagery, for instance, in roughly the way Byrne proposes we can come to know that the inner voice is speaking (2011, 117-118). In essence, when I detect an "inner picture" of an *x,* by noticing how sparse the *x* appears, I can infer that I have formed a visual image of an *x,* and thereby sensorily imagined an *x.* Of course, as in the case of judging that the inner voice is speaking, it would need to be shown that judging there to be an inner picture about *x* does not require a prior awareness of some other mental state, the knowledge of which is yet to be explained. Here my earlier remarks about the pre-conscious processes by which inner speech is distinguished from degraded acts of genuine speech perception are again relevant.

The task of explaining knowledge of our own sensory imaginings becomes much more difficult, however, if we hold that only *some* imagery-involving states count as sensory imaginings. Many will want to distinguish, for instance, episodic memories (Addis, Wong, and Schacter 2007, Addis et al. 2009), which typically involve mental imagery, from other imagery-involving processes that are less tethered to reality, using 'sensory imagination' to mark the latter (see, e.g., Kind (2001, 102)). In that case, simply knowing that one was having an "inner

picture" of an *x* would not suffice to tell one whether it was a sensory imagining of an *x*. We are then left in the same position as we are with propositional imagination: we can perhaps know that we are, in some general (imagistic) sense, thinking about an *x*, while the question of how we know that we are *sensorily imagining* an *x* remains open.

4. Why an inner sense is not the answer

So far I have said nothing about what may seem the best candidate for explaining our knowledge of our own imaginings, especially for quiet room cases. An inner-monitoring view, invoking something like an inner sense (Armstrong 1968, Lycan 1996), would provide a straightforward answer to how we know of our own imaginings. For such a "sense" would work just fine in a quiet room. Even in the absence of outward cues available to the other senses, an inner sense could take in view the contents of one's own mind and deliver a verdict. While I do not have space here for a full appraisal of inner sense views, I do wish to give a few indications of why I do not find them promising, before returning to the question of whether outward-looking or Neo-Expressivist approaches can be modified to accommodate imagination.

First, propositional imaginings are distinguished from other propositional states by their functional role. And, if only some imagery-involving states are to be counted as sensory imaginings, then the difference between sensory imaginings and (e.g.) episodic memories will likely also be a difference in functional role. Functional role is a relational property of a state, to be defined in terms of the typical causes and effects of a state with respect to other mental states and inputs and outputs to the cognitive system. As others have remarked (e.g., Goldman, 2006, 248-9), inner sense views are not well-suited to explaining our knowledge of relational features of states. This is because perceptual faculties are not, in general, the kinds of things that give us knowledge of relational or dispositional features (even if we often infer relational or dispositional features from the information about categorical features they provide). Thus, neither should an internal sense.

Goldman, in defending an inner sense view against this sort of objection, proposes that different kinds of propositional attitudes (e.g., belief and desire) may nevertheless have distinctive neurological properties corresponding to their distinct functional roles. If they did, these neurological properties—presumably categorical, intrinsic properties of the states—would

be the right sort of thing to be detected by an inner sense. And, on their basis, one could swiftly infer that the state was of a particular functional kind.

However, the idea that there really are such neural signatures corresponding to each of the distinct propositional attitude types is entirely speculative. (Goldman grants that more specifics concerning their nature "would undoubtedly be helpful" (2006, 253)). There is no consensus over whether any exist, nor any well worked-out proposals. And, even among those who believe in the robust reality of the propositional attitudes, a standard view is that there will be no type-type identities to be found between kinds described at the folk-psychological level and kinds discovered by the biological sciences (Fodor 1974).

These points have direct relevance to Nichols & Stich's (2003) Monitoring Mechanism view of introspection, which they see as a version of an inner sense view with "cognitive science trappings" (161, fn. 9). Nichols and Stich present a diagram of cognitive architecture where a "box" indicates a particular propositional attitude type, and where arrows between and among boxes indicate relations of causal interaction. Within this framework, they see the question of self-knowledge as very straight-forward. "To have beliefs about one's own beliefs," they propose, "all that is required is that there be a Monitoring Mechanism (MM) that, when activated, takes the representation *p* in the Belief Box as input and produces the representation *I believe that p* as output." Far from introducing a larger mystery, they think such a mechanism "would be trivial to implement" (161).

Now, if there literally were discrete boxes in the head, which contained within them physical sentences representing propositions toward which one took a particular attitude, then a simple mechanism of the kind Nichols and Stich propose could do the necessary work. But, of course, Nichols and Stich do not really hold that there are boxes in the head; their boxes are a kind of diagrammatic shorthand indicating that there exists in the mind a group of mental states with a common functional role. But, if this is all the boxes indicate, then any sense-like monitoring mechanism with the role of generating knowledge of a particular propositional attitude type must, among other things, be able to distinguish mental states of just that type from others. The supposedly simple mechanism that appends "I believe that" in front of each of one's beliefs cannot blindly reach into the one box to which it is attached and grab a proposition. It must ensure that it only appends "I believe that" (or "I imagine that") to propositions with the functional role of belief (or, for "I imagine that", with the functional role of imagination). We

are thus brought back to the question of how a mechanism that is genuinely sense-like could detect something like propositional attitude type, which requires detecting complex relational or dispositional features of a state. One can then do no better than to speculate that there is some neural signature for each propositional attitude type, which the monitoring mechanism might come to recognize.[9]

The inner sense theorist is no better off, then, than an outward-looking theorist who proposes that there is *some* outward looking rule we can follow to arrive at knowledge of our own imaginings, without stating what the rule is, or than the Neo-Expressivist who claims that there is some relevant performance-equivalence, without saying what it is. And the outward looking theorist and Neo-Expressivist are otherwise in a much better position, as the key ingredients of their views (the ordinary senses, rational inference, and language learning) are things we have reason to believe in independent of the question of self-knowledge. By contrast, the very idea of an inner sense only enters the scene when we are faced with the question of our knowledge of our own mind, and is suspect for that reason.

5. Reimagining imagination

I have certainly not canvassed every approach to self-knowledge. But I hope to have made a case that imagination poses a special challenge for theories of self-knowledge—a kind of challenge not posed by many other kinds of mental states. The problem is not that there are *no* instances where we can explain how we know that we are imagining. The problem is that there are too few. In particular, imaginings in a quiet room seem to present a barrier to theories of self-knowledge.

In this section I want to shift focus temporarily from the question of self-knowledge to reflect on some cases where we can tell with reasonable certainty that *someone else* is imagining. I will ultimately argue that they hold lessons with respect to the question of self-knowledge, because they are suggestive of what it is, cognitively, that we ordinarily take imagining to be.

---

[9] A referee questions whether a monitoring mechanism could be sensitive to *relational* properties of mental states, by tracking the causal source of the mental state (determining, e.g., that it resulted from stimulation of the sense modalities, and therefore is a perception). There are at least two *prima facie* problems with extending this idea to imagination: first, the cause of imaginings will presumably be the same as for other forms of stimulus-independent thought, leaving the inner sense unable to distinguish among them; and, second, any attempt to sharpen the account of the right kind of cause—stipulating, e.g., that the cause will be an intention *to imagine*—will presuppose a prior account of how that (relationally-defined) mental state is known, giving rise to a regress.

The following are three scenarios where both the person imagining, and someone watching the person imagining, are in a good position to know that the person imagining is imagining:

1) *Pretense*: As noted earlier, we can potentially come to know we are imagining that *p* by determining that we are pretending that *p,* and inferring from our pretending that *p* that we are imagining that *p*. If an otherwise sane adult crawls on all fours and says "*ruff ruff*!", we can infer that he is imagining that he is a dog.

2) *Fantasy*: Consider a situation where you are being led along in a fantasy or plan by someone else: The travel agent says, "Imagine wiggling your toes in the soft white sands, while courteous wait-staff in crisp white uniforms bring cocktails on bright silver platters…" "Ah yes," you say, "I can see it now…." Both you and someone witnessing this exchange can reasonably infer that you are imagining the vacation.

3) *Fiction*: Suppose you are reading *Crime and Punishment* at Starbucks. Some who spies you from across the café can reasonably judge you to be imagining the events the novel describes. You can reasonably judge that as well.

In each of these situations, you—and someone watching you—can reasonably infer that you are imagining something or other. Such inferences seem reasonable, if fallible. Asked why we thought such a person was imagining, we could marshal an answer. To suggest that we determine when *we ourselves* are imagining by noticing these outward signs may seem to get things backwards. However, my point for now is simply that a person reasonably *could*, if he wished, infer from these outward signs that he is imagining.

On the basis of *what* do we infer that people in such situations are imagining? What feature of each situation leads us to judge that an imagining is in progress? I want to propose that we follow a simple heuristic along the lines of: *If a person is doing some elaborated thinking about a situation she does not to take to be actual, or about an object she does not believe to exist, then she is imagining that situation or object*. I will call this the *simple heuristic*. The three cases above are relatively uncontroversial instances where we can infer that a person is imagining *just because* they are cases where there is good evidence available (to the outward senses) that the person is doing some elaborated thinking about a situation she does not believe to be actual. I assume that if we replace the novel with a textbook, in example three, suspicions

that the person is imagining diminish. This is because suspicious simultaneously diminish that the person is thinking about a situation she does not believe to be actual.

We may not always be able to tell when someone is doing some extended thinking about a situation he does not believe to be actual; but, when we can, the simple heuristic tells us we can infer that the person is imagining. Yet even if we follow the simple heuristic, this does not show that imagining can only occur when the scenario or object imagined is not believed to be actual. The simple heuristic is a proposal concerning the kind of evidence we look for when ascribing an imagining, not a metaphysical claim about the nature of imaginings. With this point in mind, we can distinguish between *belief-matching* imaginings, and *beyond-belief* imaginings. Belief-matching imaginings are imaginings that perfectly align with what the imaginer already believes to be the case. Someone who exclusively used the simple heuristic to ascribe imaginings would not detect any belief-matching imaginings. Beyond-belief imaginings are imaginings that represent situations or objects beyond those the imaginer believes to be actual. For an imagining to be "beyond-belief," in the intended sense, is not the same as for it to *conflict* with one's beliefs. An imagining can be beyond-belief, in my stipulative sense, simply by representing a situation that one neither believes nor disbelieves to be the case—e.g. imagining that Barack Obama is currently sitting down. Also, and importantly, entertaining many conditional beliefs can, in the intended sense, count as engaging in cognition that is beyond-belief. For instance, when I judge that, *if the coin lands heads, I will win a cookie*, I am thinking about a situation that is beyond what I believe to be actual –the coin landing heads—even if I in fact believe the conditional. Judging a conditional to be true counts as engaging in cognition that is beyond-belief, in my sense, whenever the antecedent of the conditional is not a proposition one already believes.

In the balance of this essay I will argue that questions of self (and other) knowledge with respect to beyond-belief imaginings will be easier to answer than the corresponding questions with respect to belief-matching imaginings. However, taking this route to self-knowledge of imagination requires broadening the scope of imagination beyond what many philosophers will want to accept. For it involves accepting that the simple heuristic is not too broad.

5.1 Beyond-belief imagining in a quiet room

If the simple heuristic is reliable—at least in the sense that following it will not lead one to *over*-ascribe cases of imagining—then cases where a person engages in some extended thought about a situation or object she believes not to be actual must usually be cases of (beyond-belief) imagining. Some will think this renders the simple heuristic much too broad, as there are many cases of such though (e.g. mere supposition, or wishing) that are not imagination. Having flagged that worry, I want to first say why the simple heuristic might be of use to someone in a quiet room, and hence relevant to the question of self-knowledge.

Let us return to Byrne's rule, THINK, which allows a person to know that she is engaged in thought about *x*. Earlier I noted that, even if we can judge the inner voice to have spoken about an *x*, THINK does not offer the means to distinguish imaginings from many other forms of occurrent thought about *x*. However, if we accept the reliability of the simple heuristic, a variation on THINK may enable us to distinguish beyond-belief imaginings from other kinds of mental states. The key is to add a step where, in effect, the rule-follower judges that the proposition expressed by the inner voice is false, or that the object represented by a visual image does not exist. Consider, then, the following:

IMAGINE THAT: "If the inner voice says '*p*' and not *p,* believe that I am imagining that *p.*"

According to IMAGINE THAT, we can infer that we are imagining that *p* whenever the inner voice says '*p*', while not-*p* is found to be the case. A version of the rule for (visual) sensory imagination is also available:

IMAGINE IT: "If the inner picture is of an *x,* and there is not an *x*, believe that I am imagining an *x*."

IMAGINE THAT does not allow one to know about *all* of one's beyond-belief imaginings. It does not allow one to know when one has imagined a proposition that one neither believes *nor* disbelieves (such a proposition is still beyond belief). Similarly, IMAGINE IT does not allow one to know when one has imagined an object one believes to exist. However, each rule would

at least offer a path to knowing one is imagining in *many* quiet room cases of (beyond belief) imagining.[10]

With this general strategy in mind, we can see how a comparable Neo-Expressivist approach might work. For reasons discussed above, there may not be a straightforward performance-equivalence between a statement that merely expresses an imagining and one that self-ascribes an imagining. However, there are *relevant* performance-equivalences we might find, with the simple heuristic in mind. In place of "I am thinking that *p,* but I don't really believe that *p,*" one can substitute, "I am imagining that *p,*" without rebuke. If we can learn when to say the former, we can learn when to say the latter. And, despite the air of Moorean paradox, we sometimes say things like "*p,* but not *really*" when we are pretending with young children. For instance, we might raise a paperclip in the air and say, "Look, this is an airplane!" and then add, *sotto voce*, with a wink and a nod, "but not really…." Often, situations where it is appropriate to say "*p,* but (*with a wink and a nod*) not really" are situations where it is performance-equivalent to say "I am imagining that *p.*" If we can learn how and when to make the former sort of statement, we can learn when to make the latter, which is a self-ascription of an imagining.

But a question looms for these proposals: is every instance where we engage in some extended thought about a situation or object we do not believe to be actual properly characterized as a case of imagination? The problem (developed in Section Three) with applying Byrne's THINK rule to imagination was that it did not offer the tools for distinguishing among many different modes of thought—be they supposing, wishing, wondering, suspecting, imagining, and so on. By the same token, one might object that there are many cases where we engage in extended thought about situations or objects we believe not to be actual, yet which are not instances of imagination. Examples include: wishing that *p,* hoping that *p,* wondering whether *p,* fearing that *p,* supposing that *p,* and so on.

There are two ways to respond to this challenge. One can grant the point, and try to appropriately sharpen the rule or procedure so as to exclude the cases that are not imaginings.

_____

[10] As a referee observes, IMAGINE IT does not obviously allow one to distinguish between when one is imagining an *x* as opposed to knowingly hallucinating an *x.* If sensory imaginings can occur unbidden, as many suppose (citing, for example, songs stuck in the head), it is not clear how we should understand the difference between hallucination and unbidden sensory imagining. Until that is clarified, the corresponding question about self-knowledge cannot be clarified. For now I grant the point that IMAGINE IT would not allow one to distinguish one's hallucinations from one's sensory imaginings.

Or, one can deny that the counterexamples really are counterexamples, by holding that imagination *includes within it* all of these other kinds of cognition.

I want to argue for taking the latter route. Despite the finer-grained uses of some in philosophy (e.g., Chalmers (2002), Yablo (1993)), it is arguable that the folk psychological notion of imagination does not cut any finer than the simple heuristic suggests. There are many cases of wishing, hoping, fearing, wondering, considering, and fantasizing that, pre-theoretically, we are happy to *also* call imagining. Consider the child eagerly imagining the gifts he may get for his birthday. In this instance, the same processing may reasonably be called wondering, hoping, wishing, fantasizing, imagining, guessing, and so on. There is no platitude, acceptance of which partly constitutes grasping the concepts in question, which forbids one or the other label. And while philosophers often distinguish between supposition, on the one hand, and a more detailed or involved process of imagining (Gendler 2000, Doggett and Egan 2007), such distinctions do not fall out of folk psychology. One can, of course, *posit* a particular kind of mental state, to explain some human capacity. But, in that case, the warrant for believing in such a state does not come from folk psychological platitudes or naïve introspection. Using a folk psychological term like 'imagining' to refer to the state will then be misleading at best.

An important positive reason for accepting the broad construal of imagination is that it aligns with the simple heuristic. For the simple heuristic is the sort of rule that can explain our third-person grasp on the notion of imagination, by offering plausible criteria by which we judge someone else to be imagining. So long as it is not mysterious how we are able to know that someone is engaged in some extended thought about a situation or object he does not believe to be actual, it is not mysterious how we are able to know someone else is imagining. This provides, in turn, a way of seeing how the term 'imagining' gains a stable foothold in public discourse. In cases where we would say that a person is engaged in some elaborated thinking about a situation or object she does not believe to be actual, we can also say, without rebuke, that she is imagining that situation or object.

This reasoning is, I admit, less than conclusive. Some may still insist that we each have an ability to clearly distinguish, through introspection, our imaginings from other sorts of cognition that would wrongly be deemed imaginings by the simple heuristic. And they will point out that we may well have such a capacity even if we have no idea how to *explain* it. After all, even if we had no idea how our eyes work, this would give us little reason to start denying that

we see things. In response, I think a more accurate analogy is this: if a person claimed that he could discriminate many more colors than ordinary folk, yet could provide no evidence to others of the ability (beyond his testimony), nor give any account of how he did so, we would be right to view his claim with skepticism. We could rightly take a rain check on speaking of his proprietary colors. This dispute can, however, continue, with each side saying the other is proposing the unusual conception (be it of colors, or of imagination) in need of special support. I hope to have gone some distance toward establishing that imagination as conceived by the simple heuristic should be the default view, due to its applicability to the kinds of cases listed above and the relatively clear picture it provides for how we learn to speak of imaginings in the first place (especially where such imaginings—like propositional imaginings—need not involve mental imagery).

That said, the simple heuristic may still seem much too narrow, if it is really is to be offered as the exclusive epistemological criterion by which we judge a person to be imagining. For it leaves us without the resources to detect any belief-matching imaginings.

5.2 *The problem of knowing of belief-matching imaginings, in any kind of room*

As close of an intuitive tie as imagination has to thought beyond belief, many will hold that we can engage in imaginings that perfectly match what we already believe. Here is a quick argument that we can indeed do so:

P1: We can imagine propositions we believe to be possible.

P2: Of the propositions we simply believe, we also believe them to be possible.

C: Therefore, we can imagine all or most of the propositions we believe.

Anyone with an interest in linking our knowledge of possibility to imagination will naturally want to stress the imaginability of propositions we in fact believe.[11] If I believe that Tuesday is garbage pickup day, then I should be able to imagine that Tuesday is garbage pickup day. Indeed, it would sound quite odd to say that, while I believe Tuesday is garbage pickup day, I cannot imagine that it is.

And yet this simple allowance gives rise to one of the most perplexing questions with respect to self-knowledge and imagination: how do I know when I have imagined a proposition,

---

[11] Appeals to imagination as a source of modal knowledge—or as "evidence" for beliefs about what is possible—have a long history in philosophy. Recent examples include Chalmers (2002), Kung (2010), and Yablo (1993).

or set of propositions, all of which I believe? And how, for that matter, do we know when someone else has done so? This is not, I take it, the same question as: how do I know when I have simply thought a proposition? For while we can reasonably broaden the scope of imagination to include all elaborated thought that is beyond-belief, broadening it to include *all* thought whatsoever would be going too far.   Even folk psychology, with its vague directives, does not equate thinking with imagining.

How can we know when people are imagining propositions they also believe? Sometimes a person may imagine a proposition she believes as part of an imaginative episode where she imagines many others she does not believe. For instance, a person pretending to have a tea party might, as part of that pretense, imagine that it is Sunday, while also believing it is Sunday. If we can detect that this is part of the pretense (and detect the pretense in the first place), we can potentially have evidence that a believed proposition is being imagined. The difficult cases are those where one is not, as part of the same imaginative episode, imagining *anything* that conflicts with or goes beyond one's beliefs—where everything that is imagined matches one's beliefs. How do we know when or if such imaginings are occurring?

In the case of sensory imagination, the comparable question is: how do we know when we have sensorily imagined an object in a way that we believe it to exist, as opposed to merely thinking about the object with the use of a mental image? (Here I assume that by 'sensory imagination' we mean something more specific than imagistic cognition in general—see Section Three).

As far as I can see, there are no good answers to these questions. Of course we can simply ask a person if she is imagining something she believes. But this just raises again the question of how she herself knows it. None of the accounts of self-knowledge considered above offer any guidance here. Even Byrne's account, to which I am most partial, admits of no amendment that would allow one to, in effect, discriminate belief-matching imaginings from other forms of belief-matching occurrent thought (such as reasoning, inferring, and judging in general).

And belief-matching imaginings do not simply present a challenge in quiet room cases. Unlike beyond-belief imaginings, there does not seem to be any general heuristic we can apply, from the first or third-person, to detect belief-matching imaginings. Our best shot at one would be to identify some task that requires belief-matching imaginings. If we can then determine that

the task is being accomplished, we can infer the occurrence of the required imaginings. The kinds of tasks and capacities people most often call on imagination to explain include: hypothetical reasoning, modal reasoning, action-planning, pretense, and mindreading. The problem is that the cases where we can reliably judge someone to have engaged in hypothetical reasoning, modal reasoning, action-planning, pretense, or mindreading are cases where such reasoning extends beyond their own existing beliefs—where they are, in some sense, thinking of things being a way that they do not already believe the world to be. These will be cases of beyond-belief imaginings.

None of this shows that we do not in fact engage in belief-matching imaginings. The puzzle is epistemological in nature, not metaphysical. Nevertheless, the difficulties we face in explaining our knowledge of such imaginings should lead us to think critically about the reasons we have for believing in them. In cases where we are inclined to say that a belief-matching imagining has occurred, we should ask: is there some alternative description of the same activity that invokes mental states for which epistemological questions are more tractable? This is a strategy I have pursued elsewhere, in arguing that propositional imagining simply involves the use of ordinary beliefs and desires (Langland-Hassan 2012).

We should also ask whether the reasons we have for believing in belief-matching imaginings are reasons that would prevent us from equating belief-matching imagining with belief-matching cognition in general. If they are not, then a theory of how we know we are engaged in a belief-matching imagining can be equated with a theory of how we know we are engaged in belief-matching cognition (and here Byrne's THINK rule may provide some answers). The real difficulty lies in providing an account of how we know we are engaged in a belief-matching imagining, where doing so is something *more specific* than simply engaging in belief matching cognition generally.

6. Conclusion

While imagination poses serious challenges to theories of self-knowledge, there is some light at the end of the tunnel. The prospects are good for an outward-looking rule that would explain how we know when we are engaged in an imagining that represents a situation beyond what we believe to be the case (or an object we do not believe to exist). Yet this is only so if we allow that more or less any elaborated cognition that represents situations or objects beyond

one's beliefs qualifies as an instance of imagination. Those who invoke a finer-grained conception of imagination still face serious obstacles to explaining how we know when we are engaged in such imagining. In addition, assuming there are cases of imagination that do not extend beyond one's beliefs, our knowledge of such imaginings remains to be explained.

Given these limitations on the present account, the progress made here may seem very modest. No doubt it is. Nevertheless, keeping in mind the kinds of conditions under which a person can plausibly learn to use a mental state term in accord with others, we should not be surprised if our knowledge of imaginings does not answer to any more narrowly defined phenomenon.

Notes

Addis, D. R., L. Pan, M. Vu, N. Laiser, and D.L. Schacter. 2009. "Constructive episodic simulation of the future and the past: Distinct subsystems of a core brain network mediate imagining and remembering." *Neuropsychologia* 47:2222-2238.

Addis, D. R., A. T. Wong, and D. L. Schacter. 2007. "Remembering the past and imagining the future: common and distinct neural substrates during event construction and elaboration." *Neuropsychologia* 45 (7):1363-77.

Armstrong, D. 1968. *A Materialist Theory of Mind*. London: Routledge.

Bar-On, D. 2004. *Speaking My Mind*. New York: Oxford Universtiy Press.

Bar-On, Dorit, and Matthew Chrisman. 2009. "Ethical neo-expressivism." In *Oxford Studies in Metaethics, vol. 4*, edited by Russ Shafer-Landau, 132-65. Oup Oxford.

Byrne, Alex. 2005. "Introspection." *Philosophical Topics* 33 (1):79-104.

Byrne, Alex. 2007. "Possibility and Imagination." *Phiosophical Perspectives* 21 (1):125-144.

Byrne, Alex. 2011. "Knowing that I am Thinking." In *Self-Knowledge*, edited by A. Hatzimoysis, 105-124. Oxford: Oxford University Press.

Byrne, Alex. 2012a. "Knowing What I See." In *Introspection and Consciousness*, edited by Declan Smithies and D. Stoljar, 183-210. Oxford: Oxford University Press.

Byrne, Alex. 2012b. "Knowing What I Want." In *Consciousness and the Self*, edited by J. Liu and J. Perry, 165-183. Oxford: Oxford University Press.

Carruthers, Peter. 2006. "Why Pretend?" In *The Architecture of Imagination*, edited by S. Nichols. Oxford: Oxford University Press.

Chalmers, David. 2002. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility*, edited by Tamar Gendler and John Hawthorne, 145-200. Oxford: Oxford University Press.

Currie, G., and I. Ravenscroft. 2002. *Recreative Minds: Imagination in Philosophy and Psychology*. Oxford: Clarendon Press.

Doggett, T., and A. Egan. 2007. "Wanting Things You Don't Want: the Case for an Imaginative Analogue of Desire." *Philosophers' Imprint* 7 (9):1-17.

Evans, Gareth. 1982. *The Varieties of Reference*. Oxford: Oxford University Press.

Fodor, Jerry A. 1974. "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)." *Synthese* 28 (2):97-115.

Gendler, Tamar. 2000. "The puzzle of imaginative resistance." *Journal of Philosophy* 97 (2):55.

Gendler, Tamar. 2005. Imagination (Addendum). In *Encyclopedia of Philosophy*, edited by Donald Borchert. Detroit: Macmillan Reference USA.

Goldman, Alvin. 2006. *Simulating Minds*. Oxford: Oxford University Press.

Harris, P., and R.D. Kavanaugh. 1993. "Young children's understanding of pretense." *Monographs of the Society for Research in Child Development* 58 (1).

Hohwy, J. 2013. The Predictive Mind. Oxford: Oxford University Press.

Kind, Amy. 2001. "Putting the Image Back in Imagination." *Philosophy and Phenomenological Research* 62 (1):85-109.

Kung, P.   . 2010. "Imagining as a guide to possibility." *Philosophy and Phenomenological Research*   81 (3):620-633.

Langland-Hassan, Peter. 2008. "Fractured Phenomenologies: Thought Insertion, Inner Speech, and the Puzzle of Extraneity." *Mind & Language* 23 (4):369-401.

Langland-Hassan, Peter. 2012. "Pretense, Imagination, and Belief: The Single Attitude Theory." *Philosophical Studies* 159:155-179.

Lillard, A. 1994. "Making sense of pretense." In *Children's early understanding of mind: Origins and development*, edited by C. Lewis and P. Mitchell, 211-234. Hove: Lawrence Erlbaum Associates.

Lillard, A., and D.C. Witherington. 2004. "Mothers' behavior modifications during pretense and their possible signal value for toddlers." *Developmental Psychology* 40 (1):95-113.

Lycan, W. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.

Miall, R.C., D.J. Weir, D.M. Wolpert, and R.C. Stein. 1993. "Is the cerebellum a Smith Predictor?" *Journal of Motor Behavior* 25:203-216.

Nichols, S., and S. Stich. 2000. "A cognitive theory of pretense." *Cognition* 74:115-147.

Nichols, S., and S. Stich. 2003. *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds*. Oxford: Oxford University Press.

Noordhof, Paul. 2002. "Imagining objects and imagining experiences." *Mind and Language* 17 (4):426-455.

Onishi, K.H., and R. Baillargeon. 2005. "Do 15-Month-Old Infants Understand False Beliefs?" *Science* 308:255-258.

Peacocke, Christopher. 1985. "Imagination, Possibility and Experience." In *Essays on Berkeley: a Teracentennial Celebration*, edited by John Foster and Howard Robinson. Oxford: Clarendon Press.

Rakoczy, Hannes, and Michael Tomasello. 2006. "Two-year-olds grasp the intentional structure of pretense acts." *Developmental Science* 9 (6):557-564.

Rakoczy, Hannes, Michael Tomasello, and T. Striano. 2004. "Young children know that trying is not pretending -- a test of the 'behaving-as-if' construal of children's early concept of 'pretense'." *Developmental Psychology* 40 (3):388-399.

Shea, N. 2014. "Reward Prediction Error Signals are Meta-Representational." *Nous* 48 (2):314-341.

Sobel, D. M. 2009. "Enabling conditions and children's understanding of pretense." *Cognition* 113 (2):177-88. doi: 10.1016/j.cognition.2009.08.002.

Sperry, R.W. 1950. "Neural basis of the spontaneous optokinetic response produced by visual inversion." *The Journal of Comparative and Physiological Psychology* 43:482-489.

von Holst, E., and H. Mittelstadt. 1950/1973. "The reafference principle." In *The Behavioral Physiology of Animals and Man. The Collected Papers of Erich von Holst*.

Wolpert, D.M., Z. Ghahramani, and M.I. Jordan. 1995. "An internal model for sensorimotor integration." *Science* 269 (5232):1880-1882.

Yablo, Stephen. 1993. "Is Conceivability and Guide to Possibility?" *Philosophy and Phenomenological Research* 53 (1):1-42.