

The Agentive Role of Inner Speech in Self-Knowledge

Abstract

Although interpretivists are right to give inner speech a central role in generating self-knowledge, they mischaracterize the precise nature of this role. Inner speech is fundamentally an action, a form of speech, and provides us with self-knowledge not by being something that we perceive (or “quasi-perceive”) and interpret, but by being something that we knowingly do. Once this is appreciated, interpretivism is undermined.

Introduction

A recently proposed account of self-knowledge that I will call “interpretivism” (Carruthers 2009, 2011; Cassam 2011, 2014) takes inner speech to play an important role in generating the kind and degree of self-knowledge that adult human beings typically have. However, proponents of this view have misunderstood the role that it plays in self-knowledge, and it is one that actually undermines interpretivism.

In line with the existing debate on the topic, by “self-knowledge” I mean coming to have knowledge about our own propositional attitude standing states (namely, our own beliefs, desires, intentions etc.) and events (namely, our thoughts, judgments, decisions etc.). I am not concerned with knowledge about oneself in a more everyday sense, of one’s appearance or geographical location, of one’s life history, or even of one’s personality, although these certainly have a claim to being called “self-knowledge”, and inner speech may also have a role to play in some or all of these.¹ I will also, for reasons that should become clear, not be concerned with knowledge of our sensory states, or even of our emotions.

I proceed as follows. I begin by introducing interpretivism and explain why inner speech is so important for the interpretivist, and show how the interpretivist

¹ See, for example, Morin 2005 on the relationship between inner speech and self-awareness.

construes it. I then look more closely at the nature of inner speech, including how and why it has developed in humans, and suggest some consequences that this has for the epistemology of inner speech. Taking this on board, I present the role that inner speech truly plays in self-knowledge, in terms of our knowledge of actions in general, and speech acts in particular. I end by reflecting on how this is in fact incompatible with interpretivism.

1. Interpretivism and Inner Speech

What I am calling “interpretivism” is the view that we only ever have unaided self-knowledge (as opposed to, say, self-knowledge from the testimony of a friend or therapist) through a process of self-interpretation.

1.1. From anti-introspectionism to interpretivism

Interpretivism is a positive thesis, built upon the negative thesis of anti-introspectionism, and, in particular, anti-introspectionism for propositional attitudes. How I know that I am visually experiencing red or feeling pain may well be a matter of introspection; what is in doubt is whether my beliefs, desires, intentions are things that I come to know by introspection. Anti-introspectionism is the claim that we *never* have introspective access to our propositional attitudes. What is meant by introspection is the following.

Introspection: Some quasi-perceptual inner sense, which purportedly provides us with direct knowledge of our own mental phenomena.

Although there are importantly different ways of thinking about introspection, what they all share is that a subject can find out about the introspectable mental phenomenon without needing to find out anything else first. To visually perceive a cup, all I need is appropriate perceptual capacities (including perhaps relevant “conceptual capacities”), and the presence of a cup (in good lighting conditions etc.). Similarly, if I have the right introspective capacities, the introspectable phenomenon can make itself known in a similar fashion. I do not, for example, need to make an inference from some other information in order to gain knowledge of that phenomenon. The anti-introspectionist (about a particular class of mental phenomena) denies this (for that class of mental phenomena). As I’ve said, the kind of anti-introspectionism that interests us here, since interpretivism is built on it, is anti-introspectionism for propositional attitudes.

Anti-introspectionism tells us how we *don’t* get knowledge of our propositional attitudes. But how *do* we typically get such knowledge? The interpretivist’s answer is: through an interpretative process. This we can gloss by quoting a prominent interpretivist:

[A]n *interpretative process* . . . is one that accesses information about the subject’s current circumstances, or the subject’s current or recent behavior, as well as any other information about the subject’s current or recent mental life. For this is the sort of information that we must rely on when attributing mental states to other people. (Carruthers 2009, p. 123, emphasis added)

1.2. Rylean symmetry

An oft-noted consequence of the interpretivist position, which is briefly mentioned in the final sentence of the passage just quoted, is the symmetry of self-knowledge and knowledge of others. Here Ryle is sometimes (e.g. Byrne 2008, Cassam 2014) quoted as a classic proponent:

The sorts of things I can find out about myself are the same as the sorts of things I can find out about other people, and the methods of finding them out are much the same. (Ryle 1949, pp.155-6)

Although this quotation is taken out of context, and an excessive focus on it yields an inaccurate portrayal of Ryle's overall position, it is clear what the message is supposed to be. Everyone agrees that I cannot introspect the mental states of others. However, the interpretivist insists, I cannot do this for *my own* mental states either, at least not my own propositional attitudes. To use terminology that Ryle wouldn't have used, for this I need to turn the same interpretative mindreading capacities I use to gain knowledge of other people's mental states upon myself.

This "Rylean symmetry" yields a couple of predictions that the interpretivist makes (see, e.g., Carruthers 2009 where these predictions are explicitly made), and which diverge from the predictions the introspectionist would make. Given that the interpretivist requires for self-knowledge, as with knowledge of others, that the prospective knower have (i) the appropriate interpretative capacity, and (ii) the right inputs for that capacity to operate upon, we get the following two predictions. The first is that there should be no dissociation between the capacity to attribute propositional attitudes to others and to oneself: somebody who is impaired in attributing propositional attitudes to others ought to exhibit impairment in self-

attribution. The second is that, even if that capacity is intact, if the relevant input is lacking (or misleading) there should be a corresponding lack of (or mistake in) self-knowledge. The first prediction seems to be born out by studies conducted on individuals with autism (Williams and Happe 2010, Williams et al. 2016). The second prediction appears to be born out by data from confabulation in a number of contexts (e.g. Gazzaniga 2000).

1.3. Why is inner speech so important to the interpretivist?

We can see the importance of inner speech to the interpretivist by reflecting on the following criticism pre-empted by Carruthers:

I seem to be able to know what I am currently thinking and planning even though I am sitting quiet and motionless... How is this possible, the critics [of interpretivism] ask, unless we have access to our own mental states that isn't interpretive, but rather introspective? (Carruthers 2009: p.123)

A central part of his response to this relies on inner speech. One type of information that clearly grounds an attribution of a mental state to someone else concerns not only a subject's behaviour, but also their utterances. If you see your friend John go to the fridge, you might think that he wants something that he believes is in the fridge. Utterances have the added feature of being more informationally fine-grained. If John says, prior to going to the fridge, "I'm getting a Coke", I can thereby refine my interpretation from "He's getting something that he believes is in the fridge" to "He's getting a Coke".

As with John, I can interpret not only my own behaviour, but also my own

utterances. Herein lies the importance of inner speech to the interpretivist. There are times when certain sources of evidence for self-interpretation (e.g. perceiving one's own behaviour, proprioceptive data etc.) are unavailable ("I am sitting quiet and motionless"), and yet we would say that we do have access to what we judge, decide, believe, desire, and so forth. As a result, there must be something else, but still sensory or quasi-sensory, which serves as the basis for our self-interpretations. Carruthers, for example, claims that data about "visual and auditory imagery (*including sentences rehearsed in inner speech*), patterns of attention, emotional feelings..." (2009, p.4, emphasis added) can also serve as the basis for our self-interpretation, and hence self-knowledge.

1.4. Inner speech as in need of interpretation

As we've just seen, interpretivists appeal to inner speech as one important form of imagery, among others, that serves as grounds for self-interpretation. More specifically, interpretivists like Carruthers and Cassam think that there is no categorical difference between interpreting your own utterances (whether in inner or outer speech) and someone else's utterances.

As Carruthers explicitly states, "[a]ll speech – whether the speech of oneself or someone else – needs to be interpreted before it can be understood" (Carruthers 2009, p.5). What this, in turn, suggests is that one needs to interpret one's utterances to ascertain both their content (e.g. disambiguate the meaning of ambiguous words), and their illocutionary force (e.g. work out if something is a question, or an assertion, or a command etc.). Here is Carruthers pre-empting an objection derived from this:

But how is it, then, that our own utterances are not ambiguous to us, in the

way that the utterances of other people often are? If I find myself thinking, “I shall walk to the bank,” then I don’t need to wonder which sort of bank is in question (a river bank, or a place where one gets money). And this fact might be taken to indicate that I must have introspective access to my intentions. However, there will generally be cues available to disambiguate our own utterances, which wouldn’t be available to help interpret the similar utterances of another. For example, just prior to the utterance I might have formed a visual image of my local bank, or I might have activated a memory image of an empty wallet. (Carruthers 2009, p.5)

I hope to show that, contrary to this view, we (typically) don’t need to interpret our own utterances (in inner or outer speech), at least not in the same way that we need to interpret the utterances of others.

To sum up, then, not only does inner speech play an important role in generating self-knowledge for the interpretivist, but it is also presented as merely one out of a number of forms of imagery that stands in need of interpretation. Granted, interpretivists do acknowledge that, at least some forms of inner speech have motoric aspects. Nevertheless, they see that as simply another way of generating the relevant imagery, the relevant “grist for the interpretative mill”. I think instead that this provides a clue to the special role that inner speech plays in the generation of self-knowledge.

To show why I think this, I’m going to address two separate but importantly related issues. On the one hand, there is the issue of what inner speech is in *homo*

sapiens, and how and why it occurs.² I will call this “the psychology of inner speech” (where “psychology” is broadly construed to include, e.g., both cognitive neuroscience and developmental psychology). Much of this interpretivists can agree with. They simply don’t take the relevant lessons from it, and that is what I move on to. That is the issue of what inner speech does for us epistemically. I will call this “the epistemology of inner speech”. I address these in turn.

2. The psychology of inner speech

The psychology of inner speech will be addressed from two perspectives. The first is in terms of how and why inner speech typically develops in humans. The second is in terms of what inner speech is, as a phenomenon, once it has fully developed.

2.1. The developmental trajectory of inner speech

How does inner speech develop, and what purpose does it serve? One very attractive theory is that inner speech starts off as overt (viz. outer) speech (Vygotsky 1987/1934). That is to say, whatever function inner speech plays, once it has developed, is played by overt speech in children who have not yet developed the capacity to engage in inner speech. This capacity to engage in inner speech is usually seen as the capacity to inhibit the overt production of speech (see Alderson-Day and Fernyhough 2014).

According to this story, inner speech is the end product of a developmental trajectory that begins with private speech. “Private speech” refers to speech that is not

² The reference to our species is to illustrate that possible cognizers could, in principle, engage in something like inner speech, but implement it in a different way.

produced for the benefit of anyone other than the speaker. Thus, although there is an important sense in which inner speech is always *de facto* private speech, pragmatics dictates that “private speech” tends to refer to *overt* private speech, rather than inner speech (since inner speech is obviously private). Young children will first, under the guidance of a caregiver, learn to reason verbally, but out loud, for the benefit of guiding their thinking and attention, and for learning to “navigate” in light of basic social norms. Over time, they learn to inhibit its overt production (which is sometimes, somewhat misleadingly, called “internalization”). Furthermore, the reason why an auditory phenomenology is often reported is because, as with any aborted overt action (motor imagery), the predictions of the sensory consequences of the would-be action come into play, activating sensory (and somatosensory) cortices (this is central to feedback, which is crucial for all successful motoric activity).³

2.2. Inner speech as a genuinely productive phenomenon

What is going on when someone is engaged in inner speech? It is tempting to think of inner speech in terms of auditory imagination. Engaging in inner speech, on such a view, consists in imagining the sound of you speaking (or imagining hearing yourself speak). This is not what inner speech is widely thought to be among the scientists who investigate it (see Alderson-Day and Fernyhough 2015, for a comprehensive review). Inner speech is a *genuinely productive* phenomenon rather than a *recreative* phenomenon. This status of being “genuinely productive” is reflected in two related claims, which I will elaborate on sequentially. The first is that inner speech is motoric. The second is that inner speech is typically (and primarily) used for making speech acts. An upshot of this is that it is unhelpful to think of inner

³ See Jeannerod 1995 for a classic presentation of this, and Adams et al. 2013 for the same within a predictive processing framework.

speech in terms of imagination.

Evidence of motoric involvement in inner speech has been empirically supported by several electromyographical (EMG) studies, measuring muscular activity during inner speech (Rapin et al. 2013) some of which date as far back as the early 1930's (e.g. Jacobsen 1931). In short, these discovered that when you engage in inner speech muscles in the face and throat, associated with speaking, are activated. This fits nicely with the Vygotskian story presented in the previous section since it suggests that motoric vestiges of the overt speech remain.

More recently, there have been brain-imaging studies presenting results that are very much in keeping with a distinction between imagined speech and a more motoric phenomenon, namely, inner speech. In particular, Tian and Poeppel (2012) and Tian, Zarati and Poeppel (2016) have shown that there are two very different ways of generating auditory-verbal imagery, namely, of activating relevant areas of auditory sensory cortices in the absence of external sensory stimulation. One, which corresponds to inner speech (which they call “articulation imagery”) is induced through “motor simulation”, i.e. is initiated “top-down” by activation in areas of prefrontal and motor cortex associated with speaking. The other, which corresponds to inner hearing/imagined speech, is induced, in line with more standard accounts of imagery (including in other modalities, such as vision) via a memory-attention-based mechanism (e.g. Kosslyn 1994), namely, the bottom-up re-creation of a sensory encounter.

As for the second point, we can distinguish (as Roessler 2016 does) between a “mere act of inner speech” and an “inner speech act”, in a way that perfectly mirrors the distinction between a “mere act of speech” and a “speech act”. There are several different ways of thinking about speech acts, and it is important to be clear about what

I mean here. To simplify somewhat, one can think of speech acts in broadly two ways: as fundamentally conventional or as fundamentally expressive. In the first sense, a “speech act” involves the bringing about of a certain state of affairs (often, but not always an institutional state of affairs) through speech, and is heavily tied to convention. This is most clearly seen in explicitly performative utterances like “I nominate...”, “I hereby sentence you...” etc, but can be extended to more mundane utterances like assertion. This is the original way of thinking of speech acts and is associated with Austin 1962 (and Searle 1969). An importantly different sense, which can be seen as growing out of Strawson’s (1964) critique of Austin, takes speech acts not to be determined by convention, but by virtue of the mental states that they express. A similar conception of speech acts is also found in Bach and Harnish (1979). This second sense of speech acts, as being the expression of particular mental states, is the sense that is relevant to us here.

The central point for our purposes is that *speech acts* are closely tied to the speaker’s mental state in a way that *mere acts of speech* are not. Examples will make things clearer. Reciting a poem, or quoting someone, or repeating an address so as to remember it, is an act of speech but it is not a speech act. This is, in part, because the speaker, in reciting, or quoting, or repeating, does not mean what is being said, and any potential variations in the subject’s mental states are compatible with the same act being performed (and variations in what is quoted, repeated, recited do not thereby signal similar variations in the subject’s mental states). In stark contrast, asserting, requesting, demanding, questioning are speech acts. These require, if they are sincere,

that the person performing them be in certain states of mind, and they thereby express those states of mind.⁴

This illustrates why inner speech is not helpfully thought of as imagined speech (or even less as rehearsed speech) but rather *as speech*. Consider the following:

1. Jane asserted that p
2. Jane imagined asserting that p
3. Jane asserted in inner speech that p

Whereas 3 suggests 1, 2 does not. In fact, 2 suggests not-1: merely imagining asserting typically rules out actually asserting (just like merely imagining raising your right hand rules out you actually doing so). On the other hand, an assertion in inner speech is a perfectly good instance of assertion. And insofar as 1 and 3 are both assertions, they both, if sincere, require that Jane be in a certain mental state (commonly thought to be believing that p). In a related manner, assertions that p are typically treated as evidence for the attribution of the mental states that they (if sincere) require (or express), in this case, believing that p. Thus if someone asserts, “Paris is the capital of France”, you will (other things being equal) think that they believe that Paris is the capital of France. The same applies to other kinds of speech act, and other kinds of speech act are intimately tied to other kinds of mental state. Orders and requests are tied to goals, questions are tied to desires to know, compliments are tied to positive evaluations, insults to negative evaluations, etc. And when people request, question, compliment or insult, if we take them to be sincere, we thereby take them to be in those states of mind.

⁴ Of course, an insincere assertion purports to express a state of mind that it doesn't in fact express, hence why it is misleading.

In short, inner speech, in its natural form (the form it takes when it is fulfilling the function for which it evolved and developed), is most certainly not a rehearsal: it's the real deal. An assertion in inner speech is an assertion, not a rehearsed assertion; an insult in inner speech is an insult (even though the target cannot hear it), not a rehearsed insult.

So far, there is nothing here that the interpretivist need disagree with, but it forms a crucial basis for my position in a way that it doesn't so much for theirs.

2.3. Three important clarifications

At this point I'd like to clarify three things. The first is that, while inner speech is not helpfully understood in terms of imagination, it may still involve imagery (in particular, a combination of sensory and motoric imagery). To see this, we need to get clear on the distinction between "imagination" and "imagery". Imagination is a personal-level phenomenon. Whereas people imagine, they don't "do imagery". Imagery is a qualitative *component* of any number of personal-level phenomena: when people imagine, remember, reminisce, reason, judge, hypothesize etc. these can (and probably will) all involve imagery to some degree (a point nicely made by Langland-Hassan 2015). In a similar way, imagery (both motoric and auditory) may be involved in an inner assertion. That does not, however, make the inner assertion nothing more than the imagery involved in its production, still less an act of imagination. An assertion in inner speech is not re-creating a state of affairs (for whatever purpose) it is bringing about a state of affairs: it is a genuine assertion.

The second thing to clarify is that inner speech acts, although they *may*, they need not involve auditory imagery at all. Nowhere is this more clearly seen than in the congenitally deaf population. Inner "speech" in the congenitally deaf population

involves visuo-motoric imagery of sign language (of signing oneself more than of *reading* sign – it is productive not receptive (see Atkinson 2006)). Furthermore, the extent to which inner speech in the hearing population is reported as auditory varies enormously (McCarthy-Jones and Fernyhough 2011).

The third clarificatory point worth raising is the possibility of non-linguistic expressive acts, and hence, as speech act theorists have long acknowledged, of non-linguistic “speech” acts (e.g. a rude gesture or a communicative glance). Thus there may be inner equivalents of these, namely, inner expressive acts that don’t make use of *speech per se*. The point is, rather, that inner speech is special in its capacity as a potential vehicle, and quintessential vehicle, of inner speech acts, of internal acts of expression. There may be other such vehicles, but these will lack the precision of the linguistic act, and will require the same aspects that enable inner speech acts to be expressive, namely, an agentive and genuinely productive (perhaps even motoric) element.

3. The epistemology of inner speech

Inner speech is an action, and, more specifically, speech. And, like overt speech, its primary use is in performing speech acts. So it is natural to ask: What knowledge do we have of our own speech acts (whether in inner or outer speech)? Since speech acts are a form of action, I will start by addressing our knowledge of our actions generally, and then refine this account to address our knowledge of our own speech acts.

3.1. Knowing what we are doing

According to the interpretivist, we know what we are doing in the same way that we know what someone else is doing: we interpret sensory or quasi-sensory information. This process of interpretation, it must be granted, is a rich and multifaceted phenomenon that takes into account a plethora of information, including dynamic and temporally extended contextual information. Taking this richness into account goes a long way towards allaying concerns, but I would like to show how this view is still not right.

A canonical indication that someone knows what they are doing is that they can accurately answer the question “What are you doing?” An interesting feature of the “doings” typically referred to in these questions is that they are not determined by what you are physically doing in a publicly observable sense. As Falvey (2000) puts it:

Suppose a friend stops by my house and wants to go for a walk, and I say, “I can’t; I’m making bread.” This could be true even if as I say it I’m sitting on the couch reading the newspaper - perhaps I’m waiting for the bread to rise before putting it in the oven. (p.22)

Somebody observing me at that point would have no reason to think that I was making bread. That is because my making bread is a temporally extended, serial operation, guided by the overarching intention to make bread. This in itself is not problematic for interpretivists, since they could simply say that there is a difference, not in the *type of process* whereby the belief in what I am doing is reached, but rather in the *amount of information* I have as the agent. After all, somebody who had more information, who had for example observed me make the dough, put it in the cupboard, and sit on the couch, might also be able to answer the question “What is he doing?” correctly, namely, by saying “He’s making bread.” This is true. However,

there are two differences between my knowledge of what I am doing, and that of even a long-term and well-informed observer. The first is to do with what we might call *engagement*. The second is to do with *authority*.

It would be odd to say that I did not know what I was doing until I was asked. My answer demonstrates knowledge that I already have. This knowledge that I have in acting knowingly is not something that is available to the observer: it is only available to the agent. This is because the agent is *engaged* in what they are doing, whereas the observer clearly can't be so engaged. This is nicely illustrated by the fact that we can turn the spotlight on the observer. Someone could enter the room and catch the observer with their face to the telescope and ask "What are you doing?" The observer might say, "I'm watching SW make bread". The observer was (in the good case) watching me knowingly. They knew that they were watching me before being asked, since that is what *they* were knowingly engaged in. I, the agent, when asked what I am doing, say "I'm making bread", not (as it would if my epistemic situation were equivalent to that of an observer) "I'm watching myself make bread".

The second difference, which is related, comes from the agent's *authority*, in knowing what they are doing, in either correcting their behavior to be in line with what they intend to be doing, or to change their mind about what they are doing. To paraphrase another nice example from Falvey (2000, p.28), suppose my neighbour is observing me through a telescope. She sees me put the kettle on and take down the teapot. She then sees me take down a tin marked "Darjeeling", and begin to spoon some of its contents into the pot. At that point she would understandably conclude that I am making Darjeeling. But suppose that my intention was actually to make Irish Breakfast tea, and I was just being absent-minded. I don't thereby update my belief

and conclude that I'm actually making Darjeeling. The knowledge of my intention to make Irish Breakfast trumps the fact that it looks like I'm making Darjeeling.

Again, this might not in itself trouble the interpretivist: the knowledge of that intention may come from my recollection that, at some point prior to making tea I said to myself (perhaps, but not necessarily, in inner speech), "Hmm... I feel like tea. I had Darjeeling this morning and I fancy a change. I'll have some Irish Breakfast." Suppose this were said out loud, an observer might also say "Oh no! He's making Darjeeling by mistake!" So far, so good. However, suppose I notice my mistake but instead say to myself, "I can't be bothered to put this tea back in the tin, and I don't feel *that* strongly about what kind of tea I have," and with a shrug I change my plans and make Darjeeling. Again, the interpretivist might say that an observer with access to all of that too might say "Oh he's changed his mind and is making Darjeeling now". But the observer is not in the same position as me, because it is *up to me* to change my mind. And I do not need to observe that I have changed my mind in order to know that I have done so. As Moran (2001) (and several others) are keen to emphasize, the observer is always one step behind the agent in this respect. These decisions are the agent's to make.

3.2. Knowing what we are saying

The interpretivist will claim that we know what we are saying in a way that is in principle no different to how we know what others are saying. As with knowing what we are doing, I don't think that this symmetry holds. Again, I appeal to differences in *engagement* and *authority*.

Contrary to what both interpretivists seem to suggest, if I'm in the middle of a debate and I make an assertive statement "p", I don't need to listen to myself to know

that my statement was an assertion as opposed to a question, or that it was an assertion that p as opposed to an assertion that q. That is because I am engaged in an activity, and my assertion that p is born out of my desire and intention to make the point that p. For example, my interlocutor might say something that I disagree with, and I engage with that directly. It would be totally mischaracterizing the situation to say that I said something, heard myself say it, and interpreted myself.

Similarly, if I say the *wrong* thing, in the sense of not saying what I mean, if I notice it at all, I have the authority to correct it. And if I don't notice it, my assumption is that I said what I meant to say (indeed that is what not noticing it amounts to). If I taste a soup that I'm cooking with a friend and I say "It needs more salt" when what I mean to say is that it needs more pepper (which I mean to say because that's what I think the soup needs), I'll either, if I notice my mistake, correct myself, or simply assume that I said "It needs more pepper" (since that is, after all, what the soup needs!). I would suggest that the latter is not particularly rare, and it strongly suggests that it is not what I actually say, in the observable sense, that predominantly determines what I think I said. Again, I am not saying that this is because we have direct introspective access to our communicative intentions (and propositional attitudes more generally), but rather that our speech acts are contextually embedded in the flow of our – predominantly world-directed, rather than self-directed – lives as agents. The overarching mistake in the interpretivist picture is that it mischaracterizes self-knowledge as an explicitly self-directed interpretative enterprise. This is thankfully not something we often do, since it is highly disruptive. And yet we are not self-ignorant as a result of not doing this.

The fact that the contexts in which we take the observational stance towards ourselves tend to be *corrective* is highly illustrative. It shows that I don't typically use

these observations to generate self-knowledge. Rather, I am using them for a sort of *supervisory monitoring*, a verification that what I am doing or saying is in line with what I already know (firmly, if implicitly) I mean to do or say (Anscombe (1957, p. 53) makes an identical point). And unlike for the observer, this third-personal observational data isn't additional evidence that enters into *competition* with earlier evidence. Where an observer might well say to herself, "Oh perhaps he has changed his mind and wants to make Darjeeling after all", I will not be tempted to consider whether my own actions (or indeed words) betray that I have, unbeknownst to me, changed my mind.

So, explicit acts of self-observation are sometimes used, but usually to monitor and correct, to keep observable behavior in line with what we already *know* we are trying to do or say. But what makes this *knowledge* then, if not the amount and quality of evidence that grounds it? In more mainstream branches of epistemology (i.e. those concerned with the nature of knowledge more generally) it is not especially controversial to claim that knowledge can be attained in the absence of evidential grounds. For example, one can appeal to reliability (e.g. Goldman 1979), or to epistemic virtue (e.g. Sosa 2007). These options could easily be exploited here (and have been by Bar-On and Nolfi 2016). We are generally very good at knowing what we are doing, and what we are saying. To say that we are good at this because we have good evidence grounding such knowledge seems to be a gross mischaracterization of our daily lives. It is rather because we are generally good at this (and we aren't always) that we count as self-knowing agents (when we get things right).

4. The Role of Inner Speech in Self-Knowledge

Now we are in a position to address the role that inner speech plays in self-knowledge. Remember that what we are concerned with is knowledge of propositional attitude *states* and *events*. For ease, I will focus on belief as an exemplar state, and judgement (of both the theoretical and practical kind) as an exemplar event, but what I say applies to all propositional attitude states and events *mutatis mutandis*.

In good cases, we know what we are doing and saying in virtue of doing and saying those things knowingly, and not in virtue of some kind of sensory or quasi-sensory evidence. But this debate is not primarily about how we know our own actions or speech acts: it is about how we know our propositional attitudes. So what is the relationship between knowing what we are doing and saying, and knowing the relevant propositional attitudes?

The relationship between action and intention is relatively straightforward. In virtue of knowing that I am ϕ ing (or indeed, trying and failing to ϕ) I thereby know that I am intending to ϕ . Things are more complicated for speech acts. For notice that, although ϕ could take a speech act as its value (speech acts are actions, after all), this extrapolation from action to speech act is not what we are after. We are not after an account of how, for example, knowledge that I am asserting that p generates knowledge that I intend to assert that p . Rather we are after an account of how knowledge that I am asserting that p can generate knowledge that I *believe* that p . In short, we are not after knowledge of the speech act, but knowledge of the mental state that the speech act (in the good case) expresses. And it isn't simply that knowledge of assertion translates into knowledge of belief, since assertions can fail to express beliefs, whether deliberately or accidentally. What account can we give of the relationship between knowledge of assertion and knowledge of belief?

Here is my favoured account. All action is revealing. It reveals the agent's state of mind. My walking across a bridge reveals my belief that the bridge is safe to walk across. But some action is *designed* to be revealing. It is expressive. Thus a smile is plausibly designed to show joy, an exclamation of "Ouch!" is plausibly designed to show pain etc. These behaviours express joy and pain respectively. How we cash out this notion of "design" is somewhat up for grabs: we can talk about biological design, sociocultural design, and even the intentional design of an individual.⁵ It probably pays to be pluralistic about this, since then we can map different kinds of design onto different kinds of expressive behaviour.

Speech acts are a particularly sophisticated form of expressive behaviour: an assertion is designed to reveal (and therefore expresses) my belief, a question is designed to reveal my interrogative state, a promise, my committed intention to do something etc. How does *this* design work? One place to start is with the observation that speech acts require certain intentions on the part of the speaker. However, it should be noted that, the intention necessarily involved in, say, an assertion, is not the intention to express a belief for the following two reasons. First, an agent (e.g. a young child) can assert without having the concept of belief (Breheny 2006 makes a similar point). Second, assuming (as we are, for the sake of argument) anti-introspectionism and the role that inner speech is supposed to play in generating self-knowledge, how can I intend to express a belief if I don't know that I have that belief, since that is what the speech act is supposed to generate for me in the first place? No: the intention must be world-directed rather than self-directed. In the case of assertion, this is the intention to inform (or glean information, in the case of a question, or commit oneself to doing something in the case of a promise etc.) and it is a matter of

⁵ For example, I might walk across the bridge, not because I simply want to get to the other side (and believe that it is safe), but because I want to *show you* that it is safe (and I thereby show you *that I believe it to be* safe). This is walking across the bridge in an expressive way.

sociocultural norms that assertions express beliefs (questions express desires to know etc). In other words, the design we need emerges out of sociocultural norms, and our practices exploit those norms. But crucially, I don't need to know those norms explicitly in order to exploit them and be subject to them. When I intend to inform (or misinform), I assert. And my assertion (if sincere) will express a belief. But that does not mean that my intention was to express a given belief. Although it is in principle possible for an asserter to have such an intention, it is an excessively sophisticated intention as a minimal condition of assertion.

A consequence of this is that I don't need to know my belief before making the assertion that expresses it. It seems therefore plausible that often an initial assertion (whether in inner or outer speech) isn't the *result* of your knowledge of your belief, but rather *generates* it. Thus I agree with Carruthers (2009, p. 125) that sometimes your first knowledge of your belief is "via its verbal expression". However, what I mean by this (both by "via" and by "expression") is rather different. Carruthers reveals this difference in the very next sentence by claiming that "such speech, like all speech, will need to be interpreted to extract its significance" (2009, p. 125). In contrast, it seems to me deeply implausible that I interpret my own speech acts. Self-knowledge of my belief is generated by my sincere and expressively apt assertion, not because I hear that assertion and attribute a belief to myself, but rather because, if that assertion is indeed sincere and expressively apt, it in itself exemplifies my knowledge of what I believe (as Bar-On 2004 rightly points out). *That* is what I would mean by "via its verbal expression".

There is, however, an important supervisory role for interpretation; for example, ascertaining whether what we have just asserted is what we intended to assert, or is in line with a belief we actually have (*viz.* is sincere). I can only tell if

I've expressed myself badly, or have been insincere, by taking an "interpretative stance" towards myself.⁶ However, it is worth acknowledging how rare this actually is, and (more importantly) how unnecessary it is in order for an agent to count as self-knowing on a given occasion. In what sense is someone who has asserted something with utmost sincerity, precision and expressive aptitude, but failed to interpret themselves, guilty of self-ignorance? I would suggest, along with Ryle himself, that if we are tempted to call this person in any way self-ignorant or epistemically defective, then we are thinking about all of this in the wrong way.

5. Conclusion

If we see the phenomenon of inner speech for what it really is, namely, as an action that is primarily used in the performance of speech acts, then we get a rather different account of the role that it plays in self-knowledge. Inner speech is not simply imagery that needs to be interpreted like the speech of another agent. It is an action, and, as with other actions, is something that we tend to engage in knowingly. We don't need to interpret ourselves to know what we are saying (in inner or overt speech) anymore than we need to interpret ourselves in order to know what we are doing. Furthermore, what we say is often an apt and knowledgeable expression of our propositional attitude states. In short, although it seems to me plausible that we do not *ever* know our beliefs and intentions, or our judgements and decisions, by introspection, we do not *only* know them by interpretation. In short, introspection and interpretation are not the only two options.

⁶ Indeed the data from autism and confabulation seems to support interpretivism because the experiments in question generate unecological scenarios where this unusual interpretative stance needs to be adopted.

Viewed in the most general light, my concern with interpretivism, which I explored here through the lens of thinking about inner speech, is not about the importance of interpretation for self-knowledge, still less is it about anti-introspectionism. I am happy to accept both of these things. It is rather about the strong “Rylean symmetry” (which Ryle himself didn’t fully subscribe to anyway) that is taken to be a consequence of interpretivism. You don’t need to posit mysterious powers of introspection, or to play down the importance of interpretation, to claim that an agent, in virtue of being an agent, knows herself in a different way to the way an observer does, or ever can. This kind of knowledge is nicely reflected in the agentic nature of inner speech, and in the role that it plays in generating self-knowledge.

References

Adams, R., Shipp, S., & Friston, K. 2013. “Predictions not commands: Active inference in the motor system”. *Brain Structure and Function*, 218, 611–643.

Alderson-Day, B. & Fernyhough, C. 2015. “Inner speech: Development, cognitive functions, phenomenology, and neurobiology”. *Psychological Bulletin* 141(5): 931-965.

Austin, J.L., 1962. *How To Do Things With Words*, 2nd Edition, ed. J.O. Urmson and M. Sbisá. Cambridge, MA: Harvard University Press.

Anscombe, G. E. M., 1957, *Intention*, Oxford: Basil Blackwell

Atkinson, J. R. 2006. "The perceptual characteristics of voice hallucinations in deaf people: Insights into the nature of subvocal thought and sensory feedback loops". *Schizophrenia Bulletin* 32(4), 701-708

Bach, K. and Harnish, R. 1979. *Linguistic Communication and Speech Acts*, Cambridge: M.I.T. Press.

Bar-On, D., 2004, *Speaking My Mind: Expression and Self-Knowledge*, Oxford: Oxford University Press

Bar-On, D. and Nolfi, K. 2016. "Belief Self-Knowledge" *Oxford Handbooks Online*.

Bilgrami, A., 2006, *Self-Knowledge and Resentment*, Cambridge, MA: Harvard University Press.

Breheny, R. 2006. "Communication and folk psychology". *Mind and Language* 21 (1):74-107.

Byrne, A. 2008, "Knowing That I Am Thinking", in Hatzimoysis (ed.) *Self-Knowledge*, Oxford University Press

- Carruthers, P. 2009 “How we know our own minds: The relationship between mindreading and metacognition”. *Behavioral and Brain Sciences* 32 (2):121.
- Carruthers, P. 2010. “Introspection: Divided and Partly Eliminated”. *Philosophy and Phenomenological Research* 80 (1):76-111.
- Carruthers, P., 2011, *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.
- Cassam, Q. 2014, *Self-Knowledge for Humans*, Oxford University Press.
- Cassam, Q. 2011, “Knowing What You Believe”, *Proceedings of the Aristotelian Society*, 111: 1–23.
- Falvey, K., 2000, “Knowledge in Intention”, *Philosophical Studies*, 99: 21–44.
- Gazzaniga, M. 2000, “Cerebral specialization and inter-hemispheric communication: Does the corpus callosum enable the human condition?” *Brain* 123:1293–326.
- Goldman, A., 1979, “What Is Justified Belief?” in G.S. Pappas (ed.), *Justification and Knowledge*, Dordrecht: Reidel, pp. 1–25; reprinted in A.I. Goldman, *Reliabilism and Contemporary Epistemology*, New York: Oxford University Press, 2012, pp. 29–49.
- Jacobsen, E., 1931, “Electrical measurements of neuromuscular states during mental activities. VII: Imagination, recollection, and abstract thinking involving the speech

musculature” *American Journal of Physiology*, 97, pp. 200–209

Jeannerod, M. 1995. “Mental imagery in the motor cortex”. *Neuropsychologia* 33, 1419–1432.

Kosslyn, S., 1994. *Image and Brain: The Resolution of the Imagery Debate*. MIT Press.

Langland-Hassan, P. 2015. “Imaginative Attitudes”. *Philosophy and Phenomenological Research* 90 (3):664-686.

Moran, R., 2001, *Authority and Estrangement: An Essay on Self-Knowledge*, Princeton, NJ: Princeton University Press.

Morin, A. 2005. “Possible Links Between Self-Awareness and Inner Speech: Theoretical Background, Underlying Mechanisms, and Empirical Evidence”. *Journal of Consciousness Studies* 12 (4-5):115-134.

Rapin, L., Dohen, M., Polosan, M., Perrier, P., & Loevenbruck, H. 2013. “An EMG study of the lip muscles during covert auditory verbal hallucinations in schizophrenia”. *Journal of Speech, Language and Hearing Research*, 56, S1882–S1893.

Roessler, J. (2016). “Thinking, Inner Speech, and Self-Awareness”. *Review of Philosophy and Psychology*:1-17.

Ryle, G., 1949, *The Concept of Mind*, New York: Barnes and Noble.

Searle, J. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.

Sosa, E., 2007, *A Virtue Epistemology*, Oxford: Clarendon Press.

Strawson, P. 1964. "Intention and convention in speech acts" *The Philosophical Review*, 73: 439–60, reprinted in Strawson, *Logico-Linguistic Papers*, London: Methuen, 1971.

Tian, X., and Poeppel, D., 2012. "Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation." *Frontiers in Human Neuroscience*. 6:314.

Williams D., Happé F., 2010. "Representing intentions in self and other: studies of autism and typical development." *Developmental Science* 13, 307–319

Williams, D., Peng, C., & Wallace, G., 2016. "Verbal thinking and inner speech use in autism spectrum disorder." *Neuropsychology Review*, 26, 394-419

Word count: 7989 (including notes and references)