

Philosophical Issues, 23, Epistemic Agency, 2013

OBSESSIVE THOUGHTS AND INNER VOICES¹

Lucy O'Brien
University College London

1. Obsessive Thoughts

My concern in this paper is to consider the nature of obsessive thoughts with the aim of getting a clearer idea about the extent to which they are rightly identified as passive or as active.

The nature of obsessive thoughts is of independent interest, but my concern with the question is also rooted in a general concern to map the extent of mental activity, and to defend the importance and centrality of a view of self-knowledge that appeals to agency. I hold that much of our mental lives is active, and that the distinctive knowledge we have of our own minds is in many cases best explained by appealing to agency. Along with many others, I take knowledge of our actions and activities to be distinctive. We know our actions and activities in a way that we do not know anything else, and in a way that is distinctively and essentially self-conscious.

I am not going to attempt to offer a proper account of what I take an action to be, nor of how we know them as we do. But, very roughly, I take actions to be psychological events—some overt physical, some covert mental—that *are* the selections an agent makes against options, where her so selecting is her taking that selection as apt. Often, but not always, an agent's taking her selection as apt is her selecting on the basis of reasons. Further, I hold that selecting as apt gives a subject immediate self-conscious *agent's awareness* of the action that is her selecting. The claim is that once we appreciate that there are mental actions and activities, as well as physical actions and activities, we can give an explanation much of our self-knowledge that draws on an understanding of our knowledge of actions in general.²

Let us call a view that explains self-knowledge by appealing to agency, the *agency view* of self-knowledge. We can distinguish between an extreme and a

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

moderate version of the view. According to an *extreme agency view* we need to appeal to agency to explain self-knowledge because all mental phenomena self-known are active. According to a *moderate agency view* we need to appeal to agency to explain self-knowledge with respect to those mental phenomena that are active. A distinct account will be required for those that are not. Both versions of the agency view are to be contrasted with the *passivist view* according to which our explanation of self-knowledge need not appeal to agency at all: there are passive mental phenomena, and whatever explains knowledge of passive mental phenomena explains knowledge of all mental phenomena.

I am a defender of a moderate agency view. I take the moderate view for two reasons. First, because there seem to me to be mental phenomena that are entirely passive—certain forms of perception for example—knowledge of which an appeal to agency will not be able to explain.³ Second, there are passive elements of active phenomenon. In particular, an agency view of self-knowledge will have to draw, in my view, on a separate and prior view of knowledge of content.⁴ I am not a passivist because I do not think that the explanations available of our knowledge of mental content, nor of our knowledge of passive mental states such as perceptions, are sufficient to capture the nature of our self-conscious self-knowledge in active cases.

A defender of the moderate version of the agency view, unlike the passivist, and unlike the defender of the extreme agency view, faces a scope question. How much of our mental phenomena is active in nature, and how much is passive? Frankfurt sets out the thought that there are both active and passive mental phenomena very helpfully:

In our intellectual processes, we may be either active or passive. Turning one's mind in a certain direction, or deliberating systematically about a problem, are activities in which a person engages. But to some of the thoughts that occur in our minds . . . we are mere passive bystanders. Thus there are obsessional thoughts, whose provenance may be obscure and of which we cannot rid ourselves; thoughts that strike us unexpectedly out of the blue; and thoughts that run willy-nilly through our heads. The thoughts that beset us in these ways do not occur by our own active doing. It is tempting, indeed, to suggest that they are not thoughts that we think at all, but rather thoughts that we find occurring within us. This would express our sense that, although these thoughts are events in the histories of our minds, we do not participate actively in their occurrence . . . It is not incoherent, despite the air of paradox, to say that a thought that occurs in my mind may or may not be something that I think.⁵

My aim in this paper will be to begin the task of identifying the scope and nature of the active, taking my cue from the way in which Frankfurt divides up the terrain. I will not get very far. I hope, however, at least to suggest that the many of the mental occurrences Frankfurt identifies as passive are better construed as mental actions.

Frankfurt identifies three kinds of ‘mental processes’ with which we are supposed not to engage, but in relation to which we function merely as ‘bystanders’:

- (i) *Obsessive thoughts*: Obsessive thoughts are thoughts where one finds oneself constantly and unwillingly thinking something. Perhaps one forms an attachment to a friend and obsessively keep thinking, on the basis of very slim evidence ‘He loves me really’. Or perhaps one is afraid of flying and about to go on a long plane journey, and despite having good evidence to the contrary one obsessively keeps thinking ‘The plane is going to crash’.
- (ii) *Out of the blue thoughts*: There are thoughts that just pop up, like when I think in the middle of writing my paper ‘today is a Thursday’.
- (iii) *Willy-nilly thoughts*: thoughts that pass ‘willy-nilly’ through one’s mind, thoughts James characterized as ‘mere restless zigzag movement’.⁶

I am going to concentrate on obsessive thoughts this paper, and want to suggest that a proper examination of obsessive thoughts reveals them, rightly understood, to be either active phenomenon in which we engage, or as passive apprehensions of content of a kind that the agency view ought already to have acknowledged as requiring a distinct account. I will suggest obsessive thoughts are in fact heterogeneous and are naturally understood to come in at least four basic kinds, on three of which they can be understood as active.

- (i) Thoughts that are mere apprehensions of content.
- (ii) Thoughts that are commitments to how the world is.
- (iii) Thoughts that are acts of inner speaking.
- (iv) Thoughts that are products of the imagination.

2. Obsessive Judgements and Entertaining of Content

Discussions of self-knowledge often overlook the fact that there is an important difference between the task of explaining how we know that our thoughts have the contents that they do, and how we know that our thoughts are of the type that they are: how we know what attitude we have to the content of the thought. We can, following, Frege, put this as a distinction between our knowing the content of our attitudes and knowing the force of our attitudes. As I understand it the agency theory is, or at least ought to be, pretty silent on the question of how we know the *content* of our attitudes. In fact most plausible accounts of how we know that we take P to be true, will trade off assumed capacity to think P, and to know it is P, rather than Q, we are thinking.⁷ In so far as an obsessive thought involves only the

passive apprehension of content the agency theorist should be happy to say that they have little distinctive to contribute here. If thought contents such as 'the plane is going to crash' can press themselves upon a thinker just as perceptual contents might, then we might expect that our knowledge of such contents be given the same explanation, but it would be an explanation beyond the ambitions and proper jurisdiction of the agency theory.

Consider an analogy with our knowledge of our bodily movements. Suppose I know that my left arm has risen. I might think that it is very important in giving the right explanation of how I know that my left arm has risen that my account reflect whether the rising was an action of mine or not. If it was an action of mine, then my explanation of how I know it had risen might be very different from the explanation if it was not an action of mine but merely a reaction, or the result of my arm being lifted by someone else. It would not be an objection to a proposed account of how I know my actions, an account that put a lot of weight on the fact that it was an *action* that was known, to say that it does not cover the case in which my body moves passively and I know that it does. The response to such an imagined objection would be: I am not trying to explain knowledge of arm rising *per se*, but only those arm risings that are actions. Similarly, I suggest the agency theorist can press the objector from obsessive thoughts for evidence that the phenomena at issue here are attitudes of mind of a kind that were the target for explanation, and not mere entertainings of content, perhaps brought on by fear, or lust or some external stimuli.

But what if we are considering a case of obsessive thoughts that are not just involuntary entertainings of content, but are rather judgements? That is, suppose that rather than there being merely entertainings of content that repeatedly come before the mind—as perceptions or memory images might—we are thinking about obsessive thoughts that are attitudes the subject takes with respect to the truth of some matter. In these cases when a thinker obsessively thinks 'He loves me really' or 'The plane will crash' she obsessively judges 'He loves me really' or 'The plane will crash'.

The question then is whether we can take such thoughts, understood as judgements, to be active. Agency theorists have claimed that in so far as judgements are commitments made by a subject with respect to the truth of some matter they are rightly seen as acts of rational agency. They are acts of mind undertaken by a subject in the basis of her taking the world to be some way, and thus fall within the scope of the active, and so within the scope of agency theory.

But there is something right about the Frankfurt description that these are thoughts that beset us despite our own rational efforts. If they are judgements, they are judgements that do not seem to be properly under our rational control: they are recalcitrant to reason and are judgements that the subject distances herself from when she is not in the grip of them. Cassam in his discussion of such cases claims that agency theorists cannot explain our

knowledge of such thoughts (which he calls ‘passing thoughts’) because they are *passive* and explains:

Passing thoughts are passive in the sense that they are (i) not necessarily responsive to reason, and (ii) states from which one can distance or dissociate oneself. (Cassam, 2011, p.3)

Let us take the two elements separately. It seems right to think that obsessive judgements are unresponsive to reason. As Cassam puts it:

Part of what makes [obsessional thoughts] obsessional is their unresponsiveness to reason. I can find myself thinking or worrying that P even though I realize that there is no reason to think or worry that P. (Cassam, 2011, p.6)

But there are two ways in which an attitude can be unresponsive to reason. First, an attitude can be formed on a faulty basis. Take a subject who tends to judge on some faulty basis that the world is a certain way. Although the options of denying P, or questioning P are available—indeed recommended—judging P is for her an apt selection given how the world strikes her. If such a subject is able on further reflection to realise that her basis for judging is not a proper one, she could be said to have been unresponsive to reason. And note such a subject might reliably and predictably have such a tendency, and indeed know that she reliably and predictably has such a tendency. Second, an attitude can be unresponsive to reason in the sense that it is based on nothing at all; in the sense that it is not grounded in the subject taking the world to be one way or another on any basis.⁸

Consider the first kind of case, of a subject who has an attitude, formed on the basis that the world strikes her in some way, which is nevertheless unreasonable and which the subject on reflection knows to be so. If what is going on in the case of obsessional judgements is that the subject has an epistemic weakness, perhaps caused by hope or by fear, which means that she tends prematurely to judge on the basis of bad or insufficient evidence, then there is no problem for the agency theorist. Suppose it were to be found that hope and fear causes subjects to make a distinctive kind of epistemic mistake in forming beliefs about what will happen. In particular, suppose that such subjects tend to take ‘possibly P’ as grounds for ‘P’. A subject suffering from such a weakness would be very likely to make obsessional judgements of the kind we have seen. In holding it to be possible that he loves her, or possible that the plane will crash, she judges that he *does* love her and the plane *will* crash. If that is what is going on, then in judging the subject is committing herself to a view about how the world is on some basis, and that is all that is required for the agency theorist to claim that judgements are acts of a rational subject. It may be that in so judging the subject makes a commitment that she is able shortly afterwards to realise was foolish and

insufficiently grounded. On reflection she realizes that something's being possible is not sufficient for thinking it will happen, and she rehearses the probabilities of it not happening and for a moment faces the pointlessness of her hope, and the irrationality of her fear. This description of what is going on shows that it is important to make a distinction between a subject's forming a judgement on some basis that she is reliably able to come to see is insufficient or lacking in warrant, and forming a judgement on absolutely no basis *at all*, indeed, forming a judgement whilst *simultaneously* accepting it must be false. The agency theorist need not acknowledge the latter as a possibility, only the former.

In response to the claim that there are 'passive judgements' that are unresponsive to reason the agency theorist can, therefore, insist that either the recalcitrance is consistent with us taking them to be active commitments by the subject to how the world is on some basis, and are thus tractable from the point of view of the agency theory approach to judgement, or are such that it is inconsistent with them being commitments at all. We get the impression that there could be a genuine commitment to the world being thus and so only by thinking it, only by switching between cases which are mere entertainings and cases which are judgements based on unstable reasons.

The second element that was supposed to make obsessive judgements passive was that the subject can distance herself from them. She can take an attitude towards them that they do not represent her considered thinking, her best reasons, or a product of her capacity for slow thinking.⁹ Nevertheless, in so far as they really are *judgements* of hers, rather than hopes, wishes, or fears—that is, in so far as they constitute her taking the world to be a particular way at that point—she cannot be thought to be simultaneously judging that 'He loves me really' or 'The plane will crash' and thinking that she is wrong to think so. Under this construal of them, obsessive thoughts are in many ways rightly thought of as repeated snap judgements that a subject withdraws from on reflection.

I think we can get a clearer view of the challenge obsessive thoughts present to the agency theory if we look at parallel case of physical actions. We see that a parallel objection to the objection from obsessive thoughts can be made to a theorist who claims that our account of our knowledge of our physical actions needs to appeal to the fact that they are exercises of agency.

Suppose that it is not contested that our knowledge of our own physical activities and actions is *particular*. By that I mean, suppose it is agreed that the nature of our knowledge of our physical activities and actions depends importantly on the fact that it is knowledge of *our actions*, and so agreed that it is not like our knowledge of others' actions and not like our knowledge of our non-active, passive movements. Suppose it is also not contested that a decent account of our knowledge of our activities and actions is going to draw on the fact that they are in contrast to some productions of the subject, exercises of agency. Suppose further that we take utterances to be a kind

of physical action, and as such we take our account of how we know our utterances to be a special case of how we know our physical actions. In so taking it we take it to be important to our account that a subject's utterances are exercises of her agency.

Suppose, now, that in response to both the general case of physical action, and to the specific case of utterances, it is objected that the agency theorist must be wrong in thinking that there is something about the fact that they are actions that explains our knowledge of them. The agency theorist must be wrong because they do not have a way of dealing with the problem of passive actions: actions with respect to which the subject seems to be passive in the sense that they are recalcitrant or not responsive to her reasons for acting, and to be such that the subject disassociates herself from them.

Like the objection from thoughts to which we are 'bystanders' there seem to be three kinds of case the objector might have in mind.

First, there are the familiar cases of obsessive or compulsive actions: repeated hand washing, repeatedly checking the door is locked, smoking another cigarette. And in the case of utterances, we have cases where a subject seems to obsessively repeat an utterance to themselves: 'I must send that email', 'I must send that email'. There are also Torette's cases. Consider for example the rare cases of *echolalia*—where a subject repeats the words of others—and *palilalia* where she repeats her own words. In all such cases the subject can regret and distance herself from her actions and utterances, acknowledge, no sooner than she has carried out them out, that she had no purpose in acting or speaking as she did.

Second, we sometimes act in such a way that our action seems to come from nowhere: we suddenly catch something falling, or stand up and walk to the window. In the case of utterances we can suddenly exclaim on some matter and surprise ourselves in doing so. Here we also have Torette's cases in which a subject seems to make an utterance that comes from nowhere, and that is completely at odds with her purposes.

Third, we have drifting action: actions such as doodling, or finger tapping with no obvious aim, or direction, by a subject's reasons for acting.

As we are concerned with the case of obsessive thoughts let us consider what we would expect the agency theorist about our knowledge of our actions to say about the challenge that obsessive actions pose to their claim that our account of our knowledge of our actions needs to be rooted in the fact that they are precisely actions? Presumably what we would expect is for the agency theorist about knowledge of actions to start sorting cases from cases. Looking at the examples of obsessive actions raised we might expect them to identify some as not really actions at all, but rather as automatic bodily reactions or responses. Torette's cases, for example, are likely to be best understood as phonic tics, on a par with blinks and throat clearing rather than as exercises of agency. In so far as they are not selections the agent has active control over, and in so far their production by-passes the

normal processes for action production, we have no reason to treat them as actions. And if they are not actions we need not expect them to be known by the subject in the same way she knows the utterances that are expressions of her agency, and which she has active control over. In contrast, some of the examples seem to be best treated as actions, but as actions done for very localized, or bad, reasons that kick in, and are recalcitrant to the further and more reflective aims of the subject. So a subject desperate to give up smoking, who nevertheless lights a cigarette in response to her intense to desire to smoke, acts and acts taking that as the apt thing to do at that time. And she does so even though her action is unresponsive to reason and is one that she distances or disassociates herself from. In such a case we seem to have a badly motivated action, and an action known to be so by its agent, but they are actions none the less, and known in the way we know actions.

There are also cases of compulsive hand-washing, and repeated soliloquizing which we might be less confident of classifying. Perhaps they too are complex tics—physical or phonic. I think more plausible is that they, like the smoking case, are active responses to a feeling that the world must be changed in some way; washing, rather than not, is for the agent the apt thing to do. Whatever the results of the case-by-case examination what seems clear is that there is room to decide that what we have is really an action. It will be an action grounded in a way that does not cohere with the agent's more reflective values and aims, and will be an action which the agent can distance herself from, but will be an action nonetheless. Thus they might be known in the way we usually know we are washing our hands, or happily smoking a cigar, or uttering 'I must send that email'.

Now, to clarify, my reason for appealing to knowledge of actions and these seemingly obsessive actions—in relation to which we may be considered as 'bystanders'—is *not* as follows: it is *inevitable* that we be agency theorists with respect to our knowledge of our actions and so such problem cases must be tractable, and if they must be tractable in the action case, why not in the obsessive thought case also? It may be that the right account of our knowledge of our actions does not need to appeal to the fact that they are actions of an agent, and it may be that there is no robust connection between the physical action case and the thought case.

My reason for appealing to obsessive actions in the action case is a much more modest one: it is that one needs to be careful in describing the nature of the phenomena if one is to be sure that the phenomena are either relevant to, or cause problems for, an agency view of self-knowledge.

An action that is selected at the time of action as apt can nevertheless be irrational from the agent's settled point of view, and known to be so by the agent on reflection. As such it is one that the agent can distance and disassociate herself from, and which can seem to be an imposition on the agent in many ways. And we need to distinguish such a case from cases which can appear to be actions, alongside other actions, when in fact they

are not an action at all: they are mere reactions or automatic responses. Both kinds of phenomena may seem to be impositions on a subject who is in some way forced to be a 'bystander'; both are marked by their being unresponsive to reason, and being such that the subject distances themselves from them. However, the former should not be seen as a problem for the agency view, and the latter should not be seen as relevant.

So, the plea is, when presented with the phenomenon of obsessive thought we need to be careful in our descriptions to keep separate the irrationally grounded mental actions—such as the irrational judgement that 'he loves me really' or 'that the plane will crash'—from merely passive entertaining of contents or images which can appear to be similar but in fact are not. In so far as we are dealing with acts of mind in which we have a subject obsessively thinking on some basis that something is the case, her attitude, and her knowledge of her attitude, fall in the catchment area of the agency view. But there seems to be no particular reason to deny that a version of the agency view will apply to them. In so far as the subject has those contents passively pressed upon her then they are indeed out of reach of the agency theorist—but a moderate agency theorist has acknowledged the need for a distinct account of passive entertainings of comment anyway.

3. Inner Speech and Imagery

So far, when faced with the question of the extent to which obsessive thoughts are active or passive we have explored the possibility that they may occur both as mere passive entertainings of content, and also occur as obsessive, but unstable, commitments to the truth of some matter by a subject. However, these possibilities do not, it seems to me, cover the possibilities. When we introspect on our obsessive thoughts there are two very different things we find ourselves doing that have not been captured by the possibilities identified so far. First, and I think this is actually what is going on a lot, if not most, of the time when we are obsessively thinking, we seem to be silently talking to ourselves. We rehearse to ourselves in inner speech certain possibilities in response to our hopes and our fears. We repeatedly and, as Ryle put it, silently soliloquize: 'he loves me really' or 'the plane is going to crash'. Second, we can find ourselves conjuring up in imagination various images that in some way represent the possibilities obsessed over. Perhaps in obsessively thinking 'he loves me really' I imagine the sweet declaration, or in thinking 'the plane is going to crash' I imagine the sharply tilting plane and the screams.

My main interest for the remainder is to consider the possibility that much, if not most, obsessive thinking is silent soliloquizing. I will suggest that, so understood, obsessive thoughts provide relatively easy cases for the agency view of self-knowledge. So understood our knowledge of our

obsessive thoughts will be subject to pretty much the same account as that which explains our knowledge of our obsessive overt speech.

Recently Cassam and Byrne, have appealed to Ryle's idea that much thinking is revealed through a kind of 'silent soliloquy'.¹⁰ Their views are very different but both take much of our thinking to be manifested by, or more tightly, constituted or disclosed by, inner speech and both suggest that we can know our thinking by being aware of such inner speech. Further, they seem to take the view that if our thoughts are revealed in inner speech our access to them is quasi-perceptual, and via some perceptual, in particular, auditory imagery produced by inner speaking. I will suggest that we ought rather to see our knowledge of inner speech as absolutely parallel to our knowledge of our outer speech. To that extent, I will suggest, we should not consider it to be known via auditory phenomena in any fundamental sense: it is rather known by an agent's awareness of what they are doing.

Much of our conscious lives seem to be taken up with inner speech, and the experimental literature bears out the view. According to Vicente & Martinez-Manrique:

Some studies show . . . inner speech takes an average of 26% of our conscious waking life, although it can vary widely from subject to subject. Other studies raise this rate to 75% or higher. (Vicente & Martinez-Manrique, 2011, p.211)

I agree with the suggestion that thinking about such speech may enable us to explain at least some of our access to our non-committal thinking. In fact, in so far as we can also judge, and question and deny, and so on, in inner speech a consideration of inner speech can play a role in our account of our awareness of some of our committal thinking also. So, I too think Ryle was on to something. However, it seems to me that we have to be careful in how we understand inner speech if it is to be able to play the role of enabling us to know our thinking. To the extent that they have an account of how we should understand the nature of inner speech and our access to it, it seems to be that have precisely the wrong one if it is to help us here.

When we turn to look at the nature of inner speech we face a distinction between what MacKay (1992) calls the 'generative component' of inner speech and the 'auditory component' of inner speech. Ryle's discussion invokes both aspects. Ryle talks about catching 'myself engaged in a piece of silent soliloquy' (1949, p.160). This suggests an appeal to the generative component of inner speech: I am aware of myself by being engaged in soliloquizing—not by being an audience to my soliloquies. On the other hand Ryle remarks that 'we eavesdrop on our own voiced utterances and our own silent monologues' (1949, p.184) suggesting that our access to our soliloquizing is as an auditor.

Ryle's descriptions suggest he is operating with a traditional, and now thought to be implausible, approach to awareness of speech. The traditional picture operates with the 'double agent assumption'. The 'double agent assumption' is the assumption that 'there is an internal speaker who speaks and an internal listener who listens' (MacKay, 1992, p.137). Something like the double agent assumption was central to traditional approaches to outer speech: there is a speaker who speaks, and a hearer who hears what the speaker is saying. The idea then was that a speaker knew what they were saying by hearing what they were saying.

Cassam and Byrne both steer clearly and quickly towards thinking of inner speech, in so far as it is going to help in explaining our knowledge of our thinking, as primarily an auditory phenomenon. So after suggesting that perhaps I am aware of my having a thought that P by being aware of saying to myself that P, Cassam says:

The hard question is: in what sense is one aware of saying to oneself that P if the saying is a saying in inner speech? Auditory metaphors are virtually inescapable. The sense in which one is aware of inwardly saying to oneself that P is that one 'hears' oneself saying to oneself that P. This is hearing with the mind's ear rather than the ears attached to ones skull. No doubt much more needs to be said about this form of awareness, but to deny its existence is to deny the existence of something which seems phenomenologically real. (2011, p.10)

Byrne is yet more clearly committed to the fundamentally auditory nature of inner speech: 'Inner speech involves representing an utterance, akin to the way in which utterances are represented when outer speech is heard' (2011, p.117) and 'In some sense—yet to be explained—one hears oneself thinking, with the "mind's ear". A natural idea then is that [a person] knows she is thinking about water . . . because she eavesdrops on herself uttering in the silent Cartesian theater' (2011, p.113).

Of course, there are in fact no sounds in the head to be picked up by any inner ear and Cassam and Byrne know this. The suggestion is not that they take there to be genuine audition going on when someone 'hears' inner speech. Rather they seem to take it that the job of hearing inner speech is the work of the auditory imagination. Byrne is explicit is drawing on auditory imagination and arguing that just as visual imagination bears important similarities to vision, so too auditory imagination bears important similarities to audition.

However, the appeal to auditory imagination, rather than simply to the capacity to inner speak, to explain our knowledge of inner speech is both costly and unnecessary. It is costly to suppose that every time a subject is conscious of inner speech they have to utilize their capacity for auditory imagination to produce auditory imagery as the manufactured objects of their awareness. We can best understand most inner speech, and in particular the

kind of inner speech that most plausibly immediately discloses my thinking, as constituted by conscious mental actions of a certain kind. In particular, we can regard them as conscious inner speaking, the awareness of which requires neither audition, nor imagination.¹¹

O'Shaughnessy surely gets it right when he says:

Talking to oneself does not consist in the active generating of the imagined sounds of words . . . The temptation to interpret talking to oneself as a form of imagining is very strong, but it must I think be resisted. (2009, p.168)

If talking to oneself is not the active generation of the imagined sounds of words we should not expect the nature of our awareness of inner speech to be somehow captured by imagined sounds. To suggest this is not to deny that the character of heard sounds will play an important role in determining the nature of our awareness of our inner speech. The phonological units that make up a person's speech—inner and outer—have the character they do because they are drawn in the normal case from heard sounds. But I want to deny that it is the work of the auditory imagination that is giving us access.

We can begin the job of making the case that we should understand our access to inner speech as agent's access to inner speaking by looking briefly at what we know about our access to our outer speech. In outer speech there is no problem about where the sounds are—we produce sounds when we speak. These sounds serve a dual purpose—they can be heard by others, allowing us to communicate, and they can also give a source of information to the speaker allowing her to track how successful her planned utterances are. However, while we do hear our own voices and while we do use what we hear to keep our speech on track it is pretty well established that our experience of speaking is not drawn exclusively, or indeed mainly by our hearing the noises that comes out of our mouths. A number of things speak against the idea that we experience our own speech primarily as auditors.

First, auditory capacities seem to be inhibited when subjects are speaking (Curio et al., 2000; Gunji et al., 2001). Second, people who are congenitally deaf know what they are saying even though they are unable to hear the sounds that come out of their mouths. They know what they are saying in virtue of being the person producing their utterances. Of course, the phenomenology of speaking for a congenitally deaf person will be different in character from that of a hearing person but that is not due to a difference in post-utterance input, but rather due to the nature of their access to the units that make up their speech. Further, it is clear that people who become progressively more and more deaf do not lose their capacity to know what they are saying. They may make more mistakes because of a source of feedback falling away, but they continue to know what they are saying much as they always have. Third, speakers seem to self-correct much too quickly for the correction to be plausibly based on aural feedback (Van der Merwe, 1997).

Of course, inner speech is not outer speech and so perhaps the elements that are most significant in fixing the character of our experience of outer speech is not to the point. However, if an explanation of the character and nature of our awareness of overt speech has to call on phonological representations that are not fixed by hearing the sounds I produce, then we have resources we can draw on if we have reason to give a non-auditory source for our experience of inner speech. And given that we produce no sounds when we inner speak we have ample *prima facie* reason to do that. Furthermore, if Vygotsky is right in taking inner speech—silent soliloquizing—to be a development, in particular an internalization of, overt soliloquizing as practiced by young children, then we have some positive reason to think that our account of the experience of inner speech would be consonant and continuous with our account of the experience of our overt speech. If inner speech is outer speech with the articulatory and productive components suppressed we will expect an account of the latter to provide an account of the former.

There is a wealth of experimental data on the nature of inner speech and it can point in different directions. Nevertheless, taken together the consensus seems to be that inner speech is most plausibly understood as inner speaking. I cannot hope to do a full review here but let me just mention a few suggestive results that suggest that inner speech is predominantly inner speaking, and inner voices are inner voicings.

First, experiencing inner speech is quite unlike hearing others', or indeed our own, overt speech. In particular, certain acoustic elements of overt speech are often absent from inner speech. For example loudness and pitch do not seem to be features of our experience of much of our inner speech, nor do sex or mood of speaker. We might if pushed say we hear ourselves in inner speech at 'neutral' pitch and loudness—but does one ever hear oneself sounding shrill, sad, anxious in inner speech? We can and do sometimes imagine speech like Edith Piaf or Groucho Marx—and then it will be natural to describe the voice as female or male, as having a French or American accent, as mournful or mischievous, and so on. But that does not seem to be the standard case. In such cases we may seem to be 'generating the imagined sounds of words'; but in the standard case of inner speaking we do not seem to be doing the same thing at all.¹²

Second, errors that subjects report occurring in inner speech correspond to errors of production not of hearing. We do not mishear words in inner speech, but we do make generative errors. The errors of the generative component of inner speech are widely held to be phonological, not articulatory. Articulatory errors, such as slurring, are not reported in inner speech, but tongue twisters—which involve phonological errors – reappear in inner speech in the same frequency as in overt speech (Dell, 1978; 1980). Also patients with articulatory problems, such as congenitally speechless patients (anarthric) and patients who are speech impaired (dysarthric) seem to show a normal capacity for memory tasks involving inner speech.

Third, there is no clear evidence of interference, from auditory input, to inner speech activities, which we would expect if the auditory system were fully engaged in inner speech MacKay (1992, p.127). It is not that very high levels of noise will not disrupt inner speech tasks—there may be times when it seems that ‘I cannot hear myself think’—but such levels would also disturb speech production generally.

None of the considerations above should lead us to deny that there is a role for audition in fixing the nature of inner speech. The phonological units of inner speech are common to inner speech and outer speech. What is being denied is that the imagination needs produce anything like imagined sounds in order for us to experience most inner speech as we do. The suggestion is rather that we know what we are saying through the phonological schemas that are part of the planning and command process that leads to overt speech. Understood as such we experience inner speech as we experience other actions—through agential knowledge.

If what has been said about inner speech is right, and if much obsessive thinking is in effect inner speaking, then we have reached the result that much obsessive thinking is active, and so a candidate for explanation by the agency view.

In the discussion of inner speech above I resisted that suggestion that we should take most inner speech as involving the imaginary production of images. However, I did not rule out the possibility. And that possibility takes me back to the second kind of case raised at the beginning of this section. In thinking about the kinds of things that can be happening when we have obsessive thoughts we raised cases in which we find ourselves not repeatedly saying to ourselves ‘He loves me really’ or ‘The plane is going to crash’ but rather in conjuring up in imagination various images that in some way represent the possibilities obsessed over: the sweet declaration and the tilting plane. Much obsessive thinking may be imagistic, and there seem to be two ways in which the imagination and memory can result in a subject being beset with imagery: an entirely passive process can result in a barrage of images being pressed on the subject, or the subject can set herself an imaginative task which she actively pursues. To determine what an agency theorist should say about our knowledge of the exercise and products of the perceptual imagination we would have to say a lot more about how to think about imagination. That is a task I am not going to take up here. It seems clear again, however, that there is no reason for the moderate agency view to be, antecedently, particularly concerned. They will expect there to be cases and cases: those passive and those not. The account of self-knowledge of passive perceptual states, which was acknowledged as necessary, will be expected to be extendable to perceptual memory and passive imagination. But in so far as we understand the active imagination as involving mental actions the moderate agency theorist might expect to have something useful to say.

Notes

- [The copyright line for this article was changed on 16 July 2014 after original online publication.]
1. I have presented some of this material at a number of places including the University of Warwick, La Universidad de Granada, the University of East Anglia, and at the CEU in Budapest, and am grateful to audiences for helpful comments: especially to Gareth Jones for email correspondence. Most notably this paper was presented at *Sofia* in Huatulco, Mexico in January 2012. I benefitted enormously from the conference, and am particularly grateful to Philippe Chuard and Kristoffer Ahlstrom for their written comments. More recently, discussion with both Matt Soteriou and Quassim Cassam has been invaluable. I am also extremely grateful to Baron Reed for his help as editor.
 2. See O'Brien 2008.
 3. Boyle 2009 is a key paper in developing a case for giving up the assumption of unity with respect to explanations of self-knowledge.
 4. See O'Brien 2005.
 5. See Frankfurt 1998, p.59. It should be noted that I am not, in using this useful passage, aligning myself with Frankfurt's rather demanding views of what might be required for something to be truly active.
 6. See James 1976, p.82.
 7. This is true of diverse agency theorists about self-knowledge: Moran 2001, Peacocke 2003, O'Brien 2005, 2008.
 8. This point is made in Ferraro 2003.
 9. See Kahneman 2011 for the distinction between fast and slow thinking. Fast thinking is efficient, non-reflective thinking, but it is often irrational. The subject can, in fast thinking, judge the world to be some way, on some basis that, on reflection, turns out to have misled them into thinking something false or irrational.
 10. See Ryle 1949, Ch. VI.
 11. Carruthers 2011 argues that inner speech acts as part of the inferential base on which we know our own minds. He holds, as I do, that we need to think of inner speech as involving inner speaking. However, like Byrne and Cassam, he thinks that our account of our access to inner speech needs to draw on auditory imagery. For him our access to inner speech is due to inner speaking generating pseudo-perceptual images in the auditory imagination, which we then have downstream access to much as we do other perceptual input.
 12. In fact even in the Edith Piaf and Groucho Marx case what we are likely to be doing is imitating Piaf or Marx in inner speech.

References

- Boyle, M. (2009) 'Two kinds of self-knowledge', *Philosophy and Phenomenological Research* 78(1): 133–164.
- Byrne, A. (2011) 'Knowing that I am thinking', in A Hatzymiosis (ed.), *Self-Knowledge*. Oxford: Oxford University Press.
- Campbell, J. J. (1999) 'Schizophrenia, the space of reasons and thinking as a motor process', *The Monist* 82.4: 609–625.

- Carruthers, P. (2011) *Opacity of Mind*. Oxford: Oxford University Press.
- Cassam, Q. (2011) 'Knowing what you believe', *Aristotelian Society Proceedings* 111(1): 1–23.
- Curio, G., Neuloh, G., Numminen, J., Jousmaki, V., and Hari, R. (2000) 'Speaking modifies voice-evoked activity in the human auditory cortex', *Human Brain Mapping* 9: 183–191.
- Dell, G.S. (1978) 'Slips of the mind', in M. Paradis (ed.), *The Fourth LACUS Forum*, pp. 69–75. Columbia, SCL Hornbeam Press.
- Dell, G.S. (1980) 'Phonological and lexical encoding in speech production: an analysis of naturally occurring and experimentally elicited speech errors', Doctoral dissertation, University of Toronto, Toronto, Canada.
- Ferrero, L. (2003) 'An elusive challenge to the authorship account: commentary on Lawlor's "Elusive Reasons"', *Philosophical Psychology* 16(4): 565–577.
- Frankfurt, H. (1998) 'Identity and Externality', in *The Importance of What We Care About: Philosophical Essays*, pp. 58–68. Cambridge: Cambridge University Press. Originally in A. Rorty (ed.), *The Identities of Persons*, pp. 239–51. Berkeley, CA: University of California Press, 1977.
- Frith, C.D. (1992) *The Cognitive Neuropsychology of Schizophrenia*. East Sussex, England: Erlbaum/Taylor and Francis.
- Georgieff, N., and Jeannerod, M. (1998) 'Beyond consciousness of external reality: a "who" system for consciousness of action and self-consciousness', *Consciousness and Cognition* 7: 465–77.
- Gunji, A., Hoshiyama, M., and Kakigi, R. (2001) 'Auditory response following vocalization: a magnetoencephalographic study', *Clinical Neurophysiology* 112: 514–520.
- Hunter, D. (2003) 'Is thinking an action?' *Phenomenology and the Cognitive Sciences* 2: 133–148.
- James, W. (1976) 'The experience of activity', Chapter 6, *Essays in Radical Empiricism*, Cambridge, MA: Harvard University Press.
- Jones, S.R. and Fernyhough, C. (2007) 'Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations', *Consciousness and Cognition* 16: 391–399.
- Kahneman, D. (2011) *Thinking, fast and slow*. New York: Farrar, Straus and Giroux
- Klinger, E. and Cox, W. M. (1987–1988) 'Dimensions of thought flow in everyday life', *Imagination, Cognition and Personality* 7: 105–128.
- Langdon, R., et al. (2009) 'The phenomenology of inner speech: comparison of schizophrenia patients with auditory verbal hallucinations and healthy controls', *Psychological Medicine* 39: 655–663.
- MacKay, D.G. (1992) 'Constraints on theories of inner speech', in D. Reisberg (ed.), *Auditory Imagery*, pp. 121–149. Erlbaum; Hillsdale, N.J.
- Martinez-Manrique, F. and Agustin, V. (2010) 'What The...! The Role of Inner Speech in Conscious Thought', *Journal of Consciousness Studies* 17(9–10): 141–67.
- Moran, R. (2001) *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.
- O'Brien, L. (2008) *Self-Knowing Agents*. Oxford: Oxford University Press.
- O'Brien, L. (2005) 'Self-knowledge, agency and force', *Philosophy and Phenomenological Research*, 71(3): 580–601.
- O'Shaughnessy, B. (2009) 'Trying and acting', in L. O'Brien and M. Soteriou (eds.), *Mental Actions*. Oxford: Oxford University Press.
- Peacocke, C. (2003) *Being Known*. Oxford: Oxford University Press.
- Ryle, G. (1949) *The Concept of Mind*. London: Hutchinson.
- Van der Merwe, A. (1997) 'A theoretical framework for the characterization of pathological speech sensorimotor control', in M.R. McNeil (ed.), *Clinical Management of Sensorimotor Speech Disorders*. New York: Thieme Medical Publishers.
- Vicente, A. and Martinez-Manrique, F. (2011) 'Inner speech: nature and functions', *Philosophy Compass* 6(3): 209–219.
- Vygotsky, S. (1934/1987) *Thought and Language*. Cambridge, MA: MIT Press. Original work published 1934.