

Hearing a Voice as one's own: Two Views of Inner Speech Self-Monitoring Deficits in Schizophrenia

Peter Langland-Hassan¹

Published online: 16 May 2015

© Springer Science+Business Media Dordrecht 2015

Abstract Many philosophers and psychologists have sought to explain experiences of auditory verbal hallucinations (AVHs) and “inserted thoughts” in schizophrenia in terms of a failure on the part of patients to appropriately monitor their own inner speech. These self-monitoring accounts have recently been challenged by some who argue that AVHs are better explained in terms of the spontaneous activation of auditory-verbal representations. This paper defends two kinds of self-monitoring approach against the spontaneous activation account. The defense requires first making some important clarifications concerning what is at issue in the dispute between the two forms of theory. A popular but problematic self-monitoring theory is then contrasted with two more plausible conceptions of what the relevant self-monitoring deficits involve. The first appeals to deficits in the neural mechanisms that normally filter or attenuate sensory signals that are the result of one's own actions. The second, less familiar, form of self-monitoring approach draws an important analogy between Wernicke's aphasia and AVHs in schizophrenia. This style of self-monitoring theory pursues possible connections among AVHs, inserted thoughts, and the disorganized speech characteristic formal thought disorder (FTD).

1 Introduction

Auditory verbal hallucinations are experienced by 60 to 80 % of patients with schizophrenia (Andreasen and Falum 1991; Sartorius et al. 1974). A less prevalent, but highly diagnostic, symptom of schizophrenia is the experience of having another person's thoughts “inserted” into one's own mind. This is commonly known as *thought insertion*

✉ Peter Langland-Hassan
Langland-Hassan@uc.edu

¹ Department of Philosophy, University of Cincinnati, 2700 Campus Way, 201C McMicken Hall, Cincinnati, OH 45221, USA

(Frith 1992; Schneider 1959). The relationship between auditory verbal hallucinations (AVHs) and reports of thought insertion is a matter of dispute. Some lump the symptoms together, holding that they are the same phenomenon at bottom, described in different ways (Graham and Stephens 2000; Langland-Hassan 2008). Others split the two, on the grounds that the different qualities patients ascribe to each kind of experience warrants viewing them as distinct phenomena (Frith 2012; Vicente 2014; Vosgerau and Newen 2007). For instance, AVHs are often reported as being similar to hearing someone else speak, and so may seem akin to a perceptual experience, while inserted thoughts are sometimes described as “soundless” (Junginger and Frame 1985).

This paper accepts the lumping view, as a working hypothesis. What one person calls “hearing another’s voice,” might easily be labeled by another as “having someone else’s thoughts in my head.” And what one patient describes as being somewhat like hearing someone speak might be described by another as being more like experiencing someone else’s verbal thought.¹ Similarly, what one person conceives of as a “soundless voice” may be conceived by another as “verbal thought.” Nor is there anything in the ambiguous and often conflicting reports of patients that stands in the way of such an assimilation.² We must keep in mind both the unusual nature of the phenomena being reported, for which there is no conventionalized terminology, and the wide variability in how people describe their inner experiences (Hurlburt and Schwitzgebel 2007; Schwitzgebel 2008). That said, those who are skeptical of such an assimilation can take what follows as a discussion of AVHs alone, without much loss.

The most common cognitive-scientific explanations of AVHs appeal to deficits in a patient’s ability to monitor his or her own inner speech (Ford and Mathalon 2005a, b; Frith 1992; Moseley et al. 2013). Inner speech here is understood as the phenomenon of speaking to oneself silently. We are using inner speech when, colloquially, we are *thinking in words*. The general idea behind self-monitoring approaches is that patients report hearing another’s voice because they do not realize they are “hearing” their own inner speech. While thought insertion is not always addressed as a separate symptom in such work, those who favor lumping thought insertion with AVHs can extend the

¹ This is not to suggest that *all* reports of inserted thoughts are reports of experiences that are voice-like. For instance, some patients seem to conceive of their inserted thoughts as “picture-like” (see, e.g., Mellor (1970, p. 17)). The point is rather that all or at least many of the experiences reported as AVHs may constitute a subset of the kind of experiences reported as inserted thoughts—roughly, the inserted thoughts that are voice-like.

² Nayani and David (1996) conducted a phenomenological survey of where 49 % of the 100 people reporting AVHs reported hearing the voices as coming “through air, as external stimuli”, whereas 38 % hear voices as located only “in internal space,” while 12 % reported both locations variably. In another study, Hoffman et al. (2008) found similarly ambiguous results when querying patients on the nature of the “voices” they hear: 79 % reported that the AVH ‘voice tone’ usually did not sound like their own voice, which might seem to support a view on which the experienced voices are similar to cases of hearing (setting aside the other 21 %). However, 70 % reported that their AVHs were *not* usually “louder” than their own verbal thought—which would suggest that the AVHs are more similar to cases of inner speech. Additionally 80 % said that their AVHs were not usually clearer than their own verbal thought; and 26.5 % reported that the voice seemed to come exclusively from outside the head. Laroi et al. (2012) report similarly wide-ranging phenomenological traits for AVHs in schizophrenia. The overall picture of the phenomenology of AVHs one gets from these studies is that they are in some ways very much like experiencing one’s own inner speech, but in other ways not. We should fully expect that patients might describe the overall experience differently—some holding that it is like hearing another speak, others emphasizing its similarity to cases of their own thought.

failure-of-monitoring view to explain inserted thoughts as well. After all, if inner speech is pre-theoretically equated with *thinking* (in words), a failure to recognize one's inner speech as one's own may lead a person to say that someone else's thoughts are in her mind. This point has relevance whether or not our best formal theory of cognition holds that we actually do think in inner speech.

But how could someone fail to realize that her own inner speech is, in fact, her own? The kind of monitoring that goes awry on such views is typically called *source monitoring*, *reality monitoring*, or *self-monitoring*.³ 'Self-monitoring' is the most common and perhaps most neutral of these terms. For those reasons I will use it here. Different self-monitoring views distinguish themselves from each other in their explanations of how or why it is that a person comes to incorrectly judge, or have the feeling that, her own inner speech was caused by someone else.

Wu (2012), and Wu and Cho (2013) have recently bucked the trend of positing self-monitoring deficits to explain AVHs, by arguing that the best-supported account of AVHs appeals only to the occurrence of *spontaneous auditory activity*. Their work raises a number of important challenges to self-monitoring accounts, which can only be met by better-clarifying what is involved in the self-monitoring of internally generated states, such as inner speech. Here I will respond to their criticisms of self-monitoring approaches by outlining two kinds of self-monitoring theory that can meet their objections. The first is fairly familiar, appealing to a lack of proper filtering (or attenuating, or dampening) of neural response to self-generated sensory signals. I call this the *filtering hypothesis*. This style of view has remained controversial since its most famous statement by Frith (1992).⁴ I argue here that this controversy is due in part to ambiguities in how we are to understand the mechanisms underlying such filtering, for inner speech. Refining a position I have advanced elsewhere (Langland-Hassan 2008), I propose a way of understanding such monitoring that meets the main objections one might raise, including those of Wu and Cho. A crucial point is that we need not understand the relevant filtering and attenuation as involving a process of prediction and comparison.

The second kind of self-monitoring view proposed here (in Section 4) is more novel, being based upon important analogies that can be drawn between Wernicke's aphasia and the experience of AVHs and inserted thoughts. While this view is still, at bottom, a self-monitoring view, it puts the emphasis on anomalies in the *semantics* of inner speech episodes, and not on their lack of sensory dampening or attenuation. For that reason it can be called a semantic self-monitoring view. This approach has the potential to explain abnormalities in both inner and overt speech in schizophrenia, and so to unite explanations of AVHs, inserted thoughts, and formal thought disorder (FTD).⁵

³ While these terms tend to be used interchangeably in the hallucinations literature, a referee alerts me that elsewhere they have distinct meanings. Self-monitoring views are generally based in theories of motor agency that appeal to phenomena such as corollary discharge (see Section 3 below), while the notions of source monitoring and reality monitoring are used in the memory literature to refer to processes by which the source and veridicality of an apparent memory are determined based on subjectively available features of the memory experience.

⁴ For recent criticisms of Frith-influenced self-monitoring approaches to AVHs, see Synofzik et al. (2008) and Vicente (2014).

⁵ Formal thought disorder is another common symptom of schizophrenia, diagnosed on the basis of patients exhibiting scattered, inscrutable, and semantically impoverished overt speech.

2 Self-Monitoring and the Matter of What one Intends

It might sound like a truism to say that reports of AVHs and inserted thoughts result from a failure of self-monitoring.⁶ Is it even conceivable that, while patients experiencing AVHs do indeed have auditory-verbal experiences of voices, their having and reporting them has nothing to do with a failure of self-monitoring?

Wu (2012) and Wu and Cho (2013) answer yes, arguing that our default position on the nature of AVHs should involve no appeal to failures of self-monitoring. They propose instead that AVHs result from “spontaneous activity” in areas of the brain responsible for the auditory representation of voices (Wu and Cho 2013, p. 3). The problem, they argue, is not one of self-monitoring, but of spontaneously having experiences as of voices talking, where these occur in the absence of any appropriate stimuli. Of course, self-monitoring views hold that patients have experiences as of voices talking as well. The important difference, on Wu (2012) and Wu and Cho’s (2013) view, is that the experiences occur *spontaneously* in the sense that they “are not something [the patient] intentionally brings about” (Wu 2012, p. 102). (To be intentionally brought about, according to Wu, is to be caused in the right way by one’s intentions (2012, p. 101).) By contrast, self-monitoring views, as Wu and most others construe them, require that the abnormal experiences in question are in fact the result of a person’s intention to create them (2012, p. 101–103). Why think that self-monitoring views must hold that AVHs are the result of a person’s intentions? The reasoning seems to be this: self-monitoring views claim that patients fail to recognize that they are in fact the agents responsible for certain inner speech episodes; we can only properly be said to be the agents of acts that we intend to carry out; therefore, self-monitoring views must hold that AVHs are caused by the intentions of the patient who reports them. If, instead, AVHs are simply caused by spontaneous activation, then self-monitoring views must be incorrect.

But this characterization of the commitments of self-monitoring views is not quite right. For the specific self-monitoring views that Wu and Wu & Cho criticize—which are based on models for motor control generally—are really only committed to the idea that AVHs result from a motor command of some kind (specifically, a motor command to generate speech or auditory-verbal imagery). This is despite the fact that many defenders of such views often speak loosely of the states being caused by one’s “intentions”. And we need to carefully consider the relation between something like an intention—which is a folk psychological notion—and that of a motor command issued by an action control system of some kind. Both in some sense cause behavior, yet cannot be equated. Intentions are closely linked to the notion of agentive control, which is itself tied up in normative considerations. For instance, we are only apt to praise or blame acts over which a person has agentive control. These are the acts which are said to be caused by the person’s intentions. While motor commands may also lead to behavior that we agree is intentional, in the normative and agentive sense, they also can cause behaviors that we are apt to classify as mere reflexes (such as ducking to avoid an errant Frisbee). Consider the case of Tourette’s syndrome. People with

⁶ Of course, as a referee notes, if by invoking “self-monitoring” one has in mind specific mechanisms (such as a forward model and comparator (Wolpert et al. 1998)) with which the task of monitoring one’s own actions is carried out, then the claim is not at all a truism.

Tourette's make involuntary and spasmodic bodily movements and verbal utterances, known as "tics". In some cases the tics include involuntary vocal utterances that are obscene, derogatory, and socially inappropriate. This particular symptom is known as *coprolalia*. We probably do not wish to say that the vocal utterances of those exhibiting coprolalia are intentional. Such utterances cause Tourette's sufferers considerable embarrassment and distress. Yet surely they result from motor commands to initiate speech (and *mutatis mutandis* for bodily tics). So it is no contradiction to speak of an unintended speech act that results from a motor command.⁷ Thus, even if motor commands exert a causal influence on action, they do not guarantee one has agential control over the action.

In noting the difference between an intention and a motor command, we threaten to collapse the distinction between Wu & Cho's spontaneous activation account and that of self-monitoring theorists. If an AVH occurs "spontaneously" whenever it is not caused by an intention (as is Wu's proposal (2012, p.101⁸)), then AVHs caused by motor commands to generate speech will count as spontaneous as well, so long as they are not intended. Thus, a self-monitoring theorist can hold that AVHs result from unintended motor commands to generate speech, and so defend a "spontaneous activation" account as well. And, on the face of it, this is indeed a reasonable proposal on the part of the self-monitoring theorist. AVHs and inserted thoughts may well be akin to internal vocal tics, a silent form of coprolalia.

However, even if both self-monitoring accounts and Wu & Cho's spontaneous activation account can appeal to spontaneous activation, there remain important differences between the two. The most basic question at issue is not whether AVHs are the products of intentions, but whether or not language *production* areas are implicated in the generation of AVHs. Wu & Cho's "bottom up" spontaneous activation account locates the causes of AVHs entirely within language *perception* sensory areas, while most self-monitoring accounts will at least partly implicate areas to do with language production. In that sense, self-monitoring accounts appeal to "top down" influences—insofar as motor commands are top down signals within a hierarchical cognitive architecture—while Wu & Cho's theory posits only "bottom up" causes for AVHs.

Wu & Cho offer as "proof of concept" for their bottom-up theory the fact that Penfield and Perot (1963) were able to generate AVHs in people by stimulating the temporal lobe. The temporal lobe includes the primary auditory cortex and areas responsible for language perception (such as Wernicke's area), but not regions associated with language production (such as the left inferior frontal gyrus and Broca's area). Yet, with the real debate between the two styles of theory now clearly in view, we can see that this experiment would only be proof of concept if it showed that "top down" language production areas were not *also* activated through this stimulation, perhaps via the recurrent network that is in any case known to exist between language production and language perception regions (Matsumoto et al. 2004). Moreover, it may have been that stimulation to language perception areas led speech signals that were already being generated in production areas to take on the form of AVHs. In short, this study gives no

⁷ Compare Frith et al.'s (2000) discussion of anarchic hand syndrome.

⁸ Wu (2012, p. 100–102) uses the term 'automatic' to describe actions and mental states that are not caused by one's intentions, whereas Wu and Cho (2013) opt for the term 'spontaneous' to describe mental states that are caused neither by one's intentions, nor by perception of an appropriate external stimulus. I stick with the term 'spontaneous' for ease of exposition.

reason to think that activation in language production areas was not necessary for the experience of AVHs in such patients. Only a study that effectively screens off any participation from language production areas would indeed be proof of concept for Wu & Cho's proposal.

I will have more to say about the relationship between language production and language comprehension, and its implications for Wu & Cho's theory, in Section Four. Here I simply emphasize that once we disentangle the notion of a speech-motor command from that of an intention, self-monitoring accounts have a lesser hurdle to jump, as they can appeal to unintended motor commands as the causes of AVHs and inserted thoughts. And spontaneous bottom-up activation accounts have a higher hurdle, as they must hold that long stretches of speech are somehow spontaneously represented, without the involvement of areas (such as left inferior frontal gyrus) known to be necessary for speech production in ordinary individuals. This is akin to holding that random activation of visual cortices might coincidentally result in visual experiences as of coherent strings of written text.⁹

3 Filtering and Attenuation, Without Prediction and Comparison

Despite the lesser hurdle I have claimed for self-monitoring accounts, simply appealing to the spontaneous (i.e., unintended) nature of AVHs and inserted thoughts does not by itself explain why they seem, to the people who have them, to be the product of an agency other than their own. For we likely have many passing thoughts and inner speech episodes that are not the result of an intention to have them, but which do not therefore seem to be the product of another's agency. Having a song stuck in one's head is a common example. This is why the self-monitoring theorist aims to give an account of the mechanisms whereby, ordinarily, our inner speech comes to seem under our own control *in some other sense* than that it is caused by one's intentions. For this would open the door for an explanation of how an inner speech episode that results from one's own motor command could nevertheless, through a failure of such mechanisms, seem as though it is the result of someone else's agency.

While emphasizing that their critique is meant to apply to all self-monitoring approaches, Wu & Cho single out for discussion the most popular version of such views, which posits the involvement of *forward models* and a prediction-comparison architecture thought to underlie perception and action more generally (Blakemore et al. 2002; Frith 1992; Wolpert et al. 1998). I now turn to assessing their critique, as a means to setting the stage for two conceptions of self-monitoring that withstand and move beyond their objections.

⁹ A referee correctly observes that, just as parts of visual cortex are "special" with respect to face perception (e.g., the fusiform face area), so too are parts of auditory cortex devoted to speech perception. This may seem to render less unlikely the possibility that random activation there would result in the representation of coherent speech. However, as the referee also notes, it remains extremely unlikely that random activation of these speech perception areas would result in long strings of coherent speech *without simultaneous activation of speech production areas* (such as the inferior frontal gyrus, which is not a part of auditory cortex). And it is the latter possibility to which Wu & Cho appeal.

3.1 The Comparator-Style Self-Monitoring Theory

It has long been hypothesized that during ordinary perception cognitive systems anticipate the sort of sensory input that will be received upon the fulfillment of an initiated motor command (Sperry 1950; von Holst and Mittelstadt 1950/1973). In recent years, this anticipation of self-caused sensory input has been conceived by many in terms of the cognitive system generating a prediction of the sensory input it expects, given the motor command initiated (Miall et al. 1993; Wolpert et al. 1998). This prediction is then said to be compared to actual *reafferent* sensory input. The element of cognitive architecture responsible for generating the prediction is generally called a *forward model*, and the mechanism which compares the prediction with actual sensory input the *comparator*. The motor command copy that triggers processing by the forward model is generally known as the *effeference copy*. This prediction and comparison process is thought to serve two purposes: first, it enables the organism to distinguish, in an immediate and non-inferential way, whether a particular sensory change was a result of its own movement or, instead, the movement of something in its environment. Sensory inputs that match predicted inputs are associated with self-created sensory changes. Second, it enables an organism to instantaneously register whether its desired action was carried out successfully, as the predicted/desired state can be compared to the final state at the moment that state is registered. This helps explain how organisms correct motor errors more quickly (200–300 ms) than would be possible if they had to reply upon ordinary visual or proprioceptive feedback (Miall et al. 1993; Wolpert et al. 1998).

To map this prediction and comparison process onto the issue of *feeling in control* of one's movements, we can suppose that an animal will feel in control of its ongoing action whenever there is a match at the comparator between the predicted and received input, and will not feel in control when there is a mismatch with resulting error signal. For cases of mismatch will typically occur when some outside factor has intervened in the action so as to throw it off course (or when the body itself has not, for whatever reason, correctly carried out the action). If a problem developed with this prediction and comparison architecture, one can see how an action that was in fact the result of a motor command—and which was in fact successfully carried out—could *seem* not to have been.

Interestingly, people with schizophrenia have shown deficits in the kind of error-correction associated with this control architecture (Blakemore et al. 2002; Frith 2012). For instance, in one well-known study, patients with schizophrenia reporting experiences of alien control were shown less able to correct errors in their movement of a joystick-controlled cursor, when visual feedback was not available, than controls and schizophrenic patients without such symptoms (Frith & Done 1989). Since such error correction in the absence of visual feedback has to rely upon the monitoring of motor commands, this is good evidence that this sort of monitoring is abnormal in people reporting experiences of alien control. Thus, theorists have naturally sought to extend this architecture to explain AVHs and inserted thoughts. For it seems to give us exactly what we need: a way of seeing how an action that is in fact under one's control—at least in the sense of being a result of a motor command—can come to seem as though it is not. And, further, it is a way that one can fail to be in control of one's action or mental state that is distinct from the ordinary way in which we may engage in automatic or otherwise unintended actions.

In the case of AVHs and inserted thoughts, the relevant act that comes to seem uncontrolled might be an episode of inner speech. If we think of an episode of inner speech as a kind of auditory-phonological representation that is registered in sensory cortex, errors in prediction and comparison might interfere with our appraisal of the causal source of that sensory representation. It may seem due to some outwardly perceived event and not our own motor commands. And that might be sufficient—or at any rate an important contributing factor—for making a person report that she is hearing voices, or entertaining someone else’s thoughts.

We can call this the *comparator-style* version of a self-monitoring account. This is the species of self-monitoring view that Wu & Cho take as their primary target. While I share some of Wu & Cho’s skepticism about the comparator view as applied to the case of AVHs and inserted thoughts, I do not share many of their motivations for it. First, they argue that self-monitoring theorists are committed to there being an “extra step” whereby one’s ordinary inner speech—which typically occurs in one’s own tone of voice, and addresses one with the first-person pronoun—is transformed so as to have the vocal and syntactic features of another’s speech, during AVHs.¹⁰ But, as Moseley and Wilkinson (2014) note in a response to Wu & Cho, surveys on the nature of inner speech find that the representation of other voices is a relatively common feature of ordinary inner speech, occurring in the many instances where one’s inner speech takes on a dialogical form (Alderson-Day et al. 2014; McCarthy-Jones and Fernyhough 2011). So it is not a special burden of self-monitoring views to explain how inner speech is transformed to have the vocal characteristics and syntax of another speaker. And in any case, when we engage in dialog with an imagined counterpart in inner speech, we do not ordinarily come to think that someone else is in fact controlling our inner speech. So the difficult question for all views—Wu & Cho’s included¹¹—is why *some* representations of the speech of others take on a pathological “alien” phenomenology, such that they actually seem to be under the control of someone else. This is what self-monitoring accounts primarily aim to explain, whether the hallucinated voice occurs in one’s own vocal tone and uses the first person pronoun to refer to the patient, or not. After all, while some AVHs are described as occurring in someone else’s tone of voice and being very similar to hearing another person speak, many others are not (Hoffman et al. 2008). Thus, a general explanation of the alien quality of AVHs must ultimately appeal to something other than their being similar to actually hearing someone else speak.

The more substantive debate engaged by Wu and Cho concerns the specifics of how a comparator architecture purports to explain how one’s inner speech, in whatever tone of voice it occurs, comes to seem as though it is not under one’s control. As discussed, the most general idea behind the comparator-style explanation is that, due to some neural deficit, there is a mismatch at the comparator when inner speech is generated, leading it to be registered as a sensory input not resulting from one’s own motor commands. That

¹⁰ Vicente (2014) raises a similar challenge to self-monitoring views.

¹¹ Wu & Cho propose that AVHs seem not to be under patients’ control because they are not intended *and* they have a phenomenological richness that is just like an experience of hearing someone else speak. “All that is required for the attributed externalization” writes Wu, “is that patients have an auditory experience that represents the presence of another’s voice” (2012, p. 99). Taking this proposal at face value, it would only explain the AVHs that are indeed like hearing someone else speak. As reviewed in the phenomenological studies above (fn. 1), this leaves very many cases of “hearing voices” untouched, including all that tend to be reported as inserted thoughts.

is, the comparator mismatch prevents the cognitive system from correctly determining the *source* of the inner speech episode. Wu & Cho explain why they find this explanation lacking, taking the inner speech signal to be a reafferent input signal:

There is nothing in the cancelation or suppression of reafference—essentially an error signal—that on its own says anything about whether the sensory input is self- or other-generated. Put simply: a zero error signal is not the same as a signal indicating *self*-generation, nor is a positive error signal the same as a signal indicating *other*-generation. After all, the computation of prediction error is used in other systems that have nothing to do with AVH, say movement control (2013, p. 3).

Their complaint is that the comparator architecture does not answer the critical question of how one's own inner speech comes to seem as though it was generated *by someone else*. For the error signal that is generated when there is a mismatch is not the same as “a signal indicating *other*-generation.” Why think it is not? Their answer is that similar error signals are triggered during activities that “have nothing to do with AVH,” such as movement control. In those cases, the error signal only carries the information that the action was not correctly completed, not that it was controlled by someone else.

Before responding to this criticism (in Section 3.2), there is another more serious worry that must be addressed with respect to comparator-style self-monitoring views. From the perspective of motor-control theory, inner speech should in the first instance be identified not with a reafferent or sensory input signal, but rather with the state that constitutes one's *prediction* of speech input. (Cf. Grush (2004) with respect to visual imagery and visual perception.) In that way, one's inner speech utterance, *qua* prediction, can be compared with the perceptual representation of one's own speech, registered through one's ears. Such an architecture would allow us to monitor and correct our own speech as we hear it, by comparing the prediction of our speech (generated by a forward model) with the sound of our speech registered through our ears. Pickering and Garrod (2013) defend a picture of language production along these lines, citing a range of studies in its favor (see also Oppenheim (2013)). From this perspective, inner speech—that is, the internal generation of auditory-phonological representations of words—occurs whenever we engage in ordinary speech aloud.¹² And what we ordinarily identify with inner speech—that is, with instances of saying things to ourselves, silently—are simply cases where we predict speech without in fact generating such speech, by suppressing, in effect, the last step in the motor routine.¹³ Note that this conception of inner speech remains compatible with the Vygotskian view that outer speech precedes inner speech developmentally, so long as “inner speech” in the Vygotskian sense is understood as the conscious phenomenon of speaking to oneself

¹² Here it is assumed that inner speech need not always be conscious, and is ultimately to be understood as a kind of neural process carrying auditory-phonological information.

¹³ The fact that we often engage in inner speech that has the characteristics of another person's voice (e.g., when imagining a conversation) still coheres with this picture of inner speech as overt speech with the motor element suppressed. In such cases we may be seen as suppressing overt speech that would have someone else's vocal characteristics, as when we imitate another person's accent and manner of speaking. This parallel would be challenged, however, if it turned out that our ability to imagine another's voice did not closely constrain and influence our ability to vocally mimic the voice. As far as I know, it is an open empirical question whether this constraint holds.

silently. For it could be that while the young child does indeed begin to generate inner speech (in the sense of a motor prediction) at the time she begins to speak aloud, it is only later that she becomes capable of generating such inner speech *while suppressing* her overt speech. And it is the use of inner speech while suppressing overt speech that constitutes the kind of phenomenon that most interests Vygotskians.

However, if the auditory phonological representations associated with inner speech are indeed best construed as *predictions* of sensory input, then, whenever we generate inner speech *without* simultaneously speaking aloud, we must be generating error signals to the effect that the prediction state does not match the actual auditory input. It is quite possible that this is in fact what occurs during what we ordinarily conceive of as inner speech. For it may be that inner speech episodes are globally broadcast (or otherwise made conscious) precisely because they are, from the point of view of the comparator system, error signals to which the larger cognitive system should be alerted. Yet this creates the following very serious problem for comparator-style accounts: if we then try to apply this architecture to AVHs and inserted thoughts, we cannot explain the sense in which they seem alien by saying that they occur in the absence of any adequately fulfilled prediction. For ordinary inner speech does as well.

The only alternative for defenders of comparator-style accounts of AVHs is to propose that, even during ordinary inner speech (i.e., when we are not speaking aloud), there is *both* a representation that serves as a prediction of certain auditory-phonological input, and a separately generated auditory phonological representation to which this prediction is compared, and which stands in the place of sensory “input”. One could then hold that it is a mismatch between *these* two signals that leads to reports of AVHs and inserted thoughts. But what could this second auditory-phonological “input” representation be but yet another self-generated state that, when things go well, matches the one we are calling a prediction? During ordinary inner speech we would, in effect, be generating *two* auditory-phonological representations to compare with each other. And what would be the point of that? Either we are able to reliably generate such representations, or we are not. Doing so twice over would not improve the reliability of either capacity, and would introduce the possibility of additional errors (via false positives), necessarily making the overall system *less* reliable. By contrast, there is a clear rationale for the comparison process when it involves comparing speech registered through one’s ears with predicted speech: in everyday situations, one has to distinguish one’s own speech from that of others, and to verify that the overt speech emitted through one’s mouth, and audible to others, matches one’s motor commands. There are no such pressures motivating an architecture that simply verifies that one is silently generating the auditory-phonological representations one intends.

3.2 The Filtering (or Dampening, or Attenuating) Hypothesis

In other work (Langland-Hassan 2008) I have tried to articulate an alternative to the comparator hypothesis that retains much of its spirit, while nevertheless being applicable to AVHs and inserted thoughts. This other approach—what I call the *filtering theory*—is not my invention. Many who nominally advance a comparator-style approach seem to have something more like the filtering hypothesis in mind (see, e.g., Ford and Mathalon (2005) and Jones and Fernyhough (2007a, b)). According to the filtering hypothesis, the best way to apply something *like* a comparator architecture to the phenomenon of AVHs

and inserted thoughts is to move away from the idea of prediction and comparison. Instead, some of the peculiar phenomenological features of AVHs and inserted thoughts—specifically, their seeming to derive from another's agency—can possibly be explained by a lack of attenuation, dampening, or what I call *filtering*. Filtering is essentially a kind of automatic attenuation of a self-generated sensory signal. And while this kind of attenuation is often associated with the prediction-comparison architecture already discussed, it need not be.

For instance, classical work on corollary discharge—such as that of Sperry and von Holst—did not invoke the forward model architecture, which involves predictions and comparisons. The notion of a prediction and a comparator mechanism was only later proposed by theorists such as Miall et al. (1993) and Wolpert et al. (1998), in theorizing about the control of bodily limb movements. Investigating the visual abilities of simple organisms like flies and fish, Sperry (1950, p. 488) simply posited “anticipatory adjustments” made with regard to the direction and speed of each eye movement. These were sufficient to let a fly discriminate between sensory changes caused by its own movements and by those caused by movements of things in its environment. And von Holst (1950, p. 150) argued that active perception involved triggering “positive” anticipatory signals that, when meeting with an appropriate “negative” incoming sensory signal at some early level of processing, would cancel out the positive signal. Neither theorist makes an appeal to *comparisons* taking place between two states in a similar cognitive format.

The filtering hypothesis holds that, in addition to whatever prediction and comparison architecture *may* be in place for monitoring overt speech, there are simpler auto-attenuation processes whereby motor commands—such as to move one's eyes, or to engage in speech—trigger an attenuating filter of sensory inputs that typically result from the act's being carried out. This sort of filtering need not involve any prediction and comparison. One way to view it is simply as an automatic direction of attention away from changes in sensory inputs of a (very general) sort that typically derive from the initiated action, with the function of maintaining optimal levels of attention on externally caused sensory changes.¹⁴ For instance, if all cases of looking to the right encode a certain amount of unnecessary detail (of the sort that might result in an unstable visual image), that specific amount of detail may be automatically deleted from any signal resulting from a case of looking to the right. And if one's own voice typically occurs within a certain frequency and timbre, signals registered through the ears during one's own speech may have those frequencies and timbres attenuated. This could explain, for example, the professor's ability to hear her students whispering in the back row, even while her own speech offers a far stronger auditory input to her ears. and indeed, less anecdotally, using MRI Christoffels et al. (2011) found that during speech the neural response in early auditory cortex to the sound of one's own voice *increases* parametrically as noise is introduced to the auditory feedback. In short, the greater the auditory input begins to diverge from the normal, anticipated sound of one's own voice, the greater the neural response to that input in early auditory cortex.¹⁵

¹⁴ This sort of automatic, involuntary direction of attention toward a stimulus is typically known as *attentional capture* (Grueschow et al. 2015; Sasin et al. 2015).

¹⁵ While Christoffels and colleagues interpret this finding as support for the view that speech involves matching predicted sensory input with actual sensory input, the results are equally compatible with a view on which specific auditory features characteristic of one's voice are automatically filtered, or attenuated, when a motor command has been issued to engage in speech.

Of course, as already observed, when we engage in inner speech there is no incoming sensory signal, in the ordinary sense, of the sort that would need to be attenuated. However, the generation of inner speech is known to involve activation in areas underlying speech production generally (Geva et al. 2011a; McGuire et al. 1996; Stephane et al. 2001). This makes it reasonable to suppose that any auto-attenuation signals triggered during outer speech, as a result of a motor command to generate speech, are triggered during ordinary inner speech as well, when such commands are also triggered. A disruption in this attenuating (or “filtering”) signal may result in some of the peculiar phenomenology of AVHs and inserted thoughts. For instance, in the absence of the kind of filtering normally applied to the auditory-phonological representations generated during inner speech, an abnormal amount of attention may be devoted to them. This is attention of a sort that is, in non-pathological contexts, automatically applied to the perception of others’ speech, and so becomes deeply associated with perceiving another’s speech. We have, then, the form of an answer to Wu & Cho’s question as to why the “error signal” generated due to a failure to anticipate one’s own inner speech should in this case be associated with perception of another’s agency.

Support for this proposal comes from work in other domains showing that people with schizophrenia lack the normal attenuation of sensitivity to self-created sensory stimuli. Studies by Blakemore et al. (1998, 2003), Ford and Mathalon (2005), Shergill et al. (2005), and Lindner et al. (2005) indicate that schizophrenic patients are often unable to achieve such attenuation, across multiple sense modalities (including proprioception). For instance, healthy subjects are generally unable to tickle themselves, their sensitivity to their own touch being greatly attenuated (Weiskrantz et al. 1971). This attenuation of self-generated sensation can also be observed at the physiological level, as activity in the somatosensory cortex is much reduced when tactile stimulation is self-applied (Frith 2005). Patients with schizophrenia do not show this attenuation, rating their sensitivity to self-touch as high as the same touch applied by someone else (Blakemore et al. 1998). Also, brain imaging reveals that responses to sound in auditory cortex are normally attenuated when a person is speaking. This attenuation may be directed at specific frequencies of the sort typically generated by one’s own speech. For instance, echoing Christoffels et al. (2011), Heinks-Maldonado (2006) found that neural responses in regions devoted to speech perception were reduced when people spoke and listened to their own speech on headphones, compared to when they heard a pitch-shifted distortion of their own speech. This pattern of attenuation is not found in many schizophrenic patients, particularly those who report auditory hallucinations (Ford and Mathalon 2005; Ford et al. 2001; Shergill et al. 2014). Ford and Mathalon (2005) found that, while inner speech reduces responsiveness of auditory cortex in healthy controls, schizophrenic patients lacked this attenuation during inner speech. And, in the case of vision as well, Lindner et al. (2005) found that schizophrenic patients with delusions of influence were more likely to attribute the sensory consequences of their own eye movements to the environment rather than to themselves.

An important virtue of the filtering hypothesis is that it is well-placed to explain the unusual phenomenology of *both* AVHs and inserted thoughts, while at the same time having applicability to the broader sensory-perceptual impairments found in schizophrenia. As noted above, some patients experiencing voices describe them as being like

hearing someone else speak, while others report them as being not unlike their own verbal thought. The wide variations and gradations in such reports favors an approach that locates the phenomenology somewhere between what it is like to engage in inner speech and what it is like to hear others speak. This is the kind of phenomenology to expect if one's own inner speech were suddenly (and automatically) treated by downstream processes as an auditory input, in a manner usually reserved for the speech of others.

It bears noting that, despite the fact that I have shied away from describing the relevant sort of filtering as a “prediction”, the overall theoretical picture described, and the sort of data that support it, are quite compatible with what is sometimes known as a *predictive processing framework* for understanding the computations underlying perception and sensorimotor coordination (Hohwy 2013; Wilkinson 2014). The basic idea behind such approaches is that the brain continually predicts sensory input based on its best hypotheses, seeking to reduce prediction error. The input signals that receive downstream processing—and ultimately become conscious—are primarily those that are *not predicted*. This is another way of saying that the signals that become conscious are those that are not attenuated by an appropriate filter. The main difference is that the filter hypothesis views the relevant predictions as a simple commands to attenuate certain aspects of the input, as opposed to states that are in the same format as the input and which are subsequently *compared* to the input. That difference in how the computation is implemented aside, both approaches agree that the end result is increased attention and cognitive resources automatically devoted to those aspects of sensory input that are not in some sense anticipated.

Wu objects to proposals that invoke a failure of attenuation by observing that they would have difficulty explaining cases where patients report hearing *whispering* voices (2012, fn. 12). For in such cases, what makes the mental episode seem to be under another's control is (presumably) not its greater apparent volume, compared to one's normal inner speech. However, in advocating a filtering or attenuation approach—where a lack of filtering is what leads one's inner speech to seem under another's control—one need not conceive of the filtering simply in terms of its decreasing the vivacity or represented “loudness” of the episode (though those are legitimate ways in which it may often affect the episode (Langland-Hassan 2008)). The filtering can, again, be conceived as an automatic attention shift away from certain signals or frequencies. Compare the “cocktail party effect” whereby a person engaged in a conversation may suddenly shift attention to a conversation at the other side of the room, after hearing a salient word (such as one's own name). In such cases, the person whose attention has shifted to the further off conversation may still be experiencing the person who is speaking in front of him as speaking at the same volume as before. Yet his attention is consumed with the further off (and comparatively quieter) conversation. Thus attention can be assigned to a stimulus without making it subjectively louder, and without making the signals thereby ignored subjectively quieter. On the filtering hypothesis, this sort of attention—usually reserved for the speech of others—is automatically captured by one's own inner speech, due to a lack of automatic filtering, and is usually reserved for the perceived speech of others. The idea is thus not that attention itself gives the inner speech an alien quality (we can, after all, attend to our own inner speech without making it seem to be controlled by someone else), but that, due to filtering deficits, one's attention is *automatically directed at* or *captured by* one's

unfiltered inner speech, in a way that normally only occurs when one is registering the speech of others. It is the automaticity of the attentional capture that is crucial.

Of course, many questions remain open for such an account. One of the most pressing is the intermittent nature of AVHs and inserted thoughts. It is not as though *all* of the inner speech of people suffering AVHs and inserted thoughts seems not to be under their own control. If the envisioned sort of filtering omissions occur, it seems they do so intermittently. Why would this be? Wu & Cho press this questions especially forcefully in noting that patients sometimes report being in conversation with the “voices” they hear. Could it really be that the filtering deficits occur only precisely when the “other” voice speaks? This sort of case is in fact a challenge for all accounts. (If, as Wu & Cho propose, AVHs result from spontaneous sensory activation, how is it that they occur in a way that allows for a meaningful conversational back and forth between the voices and the patient’s inner speech?) Despite such open questions, however, the filtering approach remains a promising research program well worth pursuing at this stage of inquiry.

4 Self-Monitoring and the Aphasias

I turn now to discussing the aphasias, with two goals in mind. First, reflection on Broca’s and other aphasias where language-production is impaired can help us better understand what is at stake in debates between “top-down” self-monitoring views and “bottom up” approaches like that of Wu & Cho. And, second, by considering analogies between Wernicke’s aphasia and the experience of AVHs and inserted thoughts, we can see the way to a new form of self-monitoring account, which already has empirical support.

4.1 Broca’s Aphasia and Top-down vs. Bottom-up Explanations of AVHs

People with aphasia have language deficits acquired as a result of a brain lesion, typically due to a stroke. The aphasias are specifically *language* deficits: a person who, for instance, cannot vocally produce speech but who has intact comprehension and writes sentences fluently likely does not have aphasia, but rather *verbal apraxia of speech*. While there are many sub-types of aphasia, the language difficulties faced by those who have it are generally sorted into two groups: difficulties with production, and difficulties with comprehension. It should be noted, however, that practically all people with aphasia have *some* degree of impairment in both production and comprehension (LaPointe 2005).

People with Broca’s aphasia primarily have language production difficulties. This form of aphasia is caused by damage to the left inferior frontal gyrus (for those who are left-hemisphere dominant for language), which houses what is commonly known as Broca’s area. Asked to name a collection of ordinary objects, patients with Broca’s aphasia typically respond with slow, sparse, effortful, and halting speech. These production deficits can range from relatively mild to cases where the patient can name few to no objects at all. Patients with aphasia this type typically give the impression of knowing what various presented objects are, despite being unable to come up with the proper words to name them. Nevertheless, their comprehension of

language—both spoken and written—may be more or less intact for everyday interactions (LaPointe 2005).

There is ample evidence that, just as Broca's area is needed to generate speech aloud, it is involved as well in the production of inner speech (Geva et al. 2011b; Jones and Fernyhough 2007a, b; McGuire et al. 1996). And, as self-monitoring theories would predict, Broca's and other portions of the inferior frontal gyrus are selectively activated during AVHs as well (McGuire et al. 1993; Raji and Riekkki 2012). In a recent study, Hoffman et al. (2011) found increased activation in left inferior frontal gyrus at times when participants reported (via a button press) experiencing AVHs. And, conducting a meta-analysis of many studies that compare the neural activity of people while they report experiencing AVHs to when they report the absence of AVHs, Kühn and Gallinat (2010) conclude that "the state of experiencing AVHs is primarily related to brain regions that have been implicated in speech production i.e., Broca's area" (2010, p. 779) (See also, Allen et al. (2012)).

Above, in Section 2, I noted that Tourette's syndrome may be a case where people generate involuntary speech, due to the triggering of motor speech commands in the absence of corresponding intentions. The language-production aphasias suggest the opposite dissociation: people intend to name the objects presented to them by experimenters, but, due to their lesions in language production areas, are unable. This is further support, then, for the idea that there are two kinds of top-down causes potentially relevant to explaining AVHs: intentions (however they should be understood) and speech motor commands. The latter rely upon neural networks housed (in most people) in the left inferior frontal gyrus.

Top-down activation from language production areas, such as the inferior frontal gyrus, should not be necessary for the experience of AVHs on Wu & Cho's bottom-up account. As discussed above, their theory holds that spontaneously triggered representations housed in primary auditory cortex (itself within the superior temporal gyrus) are sufficient for the experience of AVHs. And, in terms of neural geography, the primary auditory cortex and superior temporal gyrus lie well posterior to the left inferior frontal gyrus and Broca's area (though the respective regions are connected by a white matter pathway called the arcuate fasciculus¹⁶). Indeed, the considerable distance between the two regions accounts for why a stroke will often result in a lesion to one but not the other area.

In response to data showing activation of Broca's area during AVHs, Wu & Cho propose that such activation may be a result not of the AVHs themselves, but of patients' internal responses to the AVHs, which often take the form of inner speech (2013, p. 4). Whatever the merits of that response, it is important to note that, on their account, a person with severe Broca's aphasia, who cannot willfully generate speech, should nevertheless be capable of experiencing AVHs. (This is *not* to suggest that people with aphasia typically do experience AVHs, of course). This suggests an interesting prediction that would help arbitrate between theirs and top-down approaches. Suppose that a person with schizophrenia and a known history of AVHs had a stroke and, as a result, acquired severe Broca's aphasia. On top-down, self-monitoring approaches, their aphasia should lead to a comparable decrease in AVHs. On the bottom-up approach, however, we should expect the AVHs to continue

¹⁶ Lesions to the arcuate fasciculus can result in conduction aphasia, which is marked by many of the same language-production deficits as Broca's aphasia.

unabated, despite the fact that the person herself can no longer willfully generate speech. Unfortunately, I am not aware of any existing reports that confirm either prediction.¹⁷ But it is useful, in the meantime, to understand the different predictions. There is, it seems, a real dispute between self-monitoring and bottom-up approaches, which can potentially be resolved by specific dissociations.

With this in mind, investigators might try to artificially disable or suppress areas responsible for speech generation in people suffering AVHs and inserted thoughts, to see if a comparable decrease in hearing voices is achieved. For instance, Aziz-Zadeh et al. (2005) were able to suppress inner speech by applying transcranial magnetic stimulation (TMS) to the left inferior frontal gyrus (which houses Broca's area). This suggests that such stimulation, if applied to people experiencing AVHs, might help suppress AVHs as well. If, on the other hand, TMS were applied to the inferior frontal gyrus at moments when a patient reported having an AVH *without* disrupting the AVH, this would be fairly strong evidence against a role for language production regions in the genesis of AVHs. Another possibility worth considering while pursuing such results, however, is that it may be the *right* homologue of Broca's area—what is, essentially, the language production area of the non-dominant hemisphere—that is responsible for generating AVHs. There is, in fact, imaging data supporting this interesting hypothesis (Sommer and Diederer 2009).

In any event, as technologies improve for safely suppressing activity in areas necessary for speech production, our ability to assess self-monitoring accounts against bottom-up approaches will improve as well.

4.2 Wernicke's Aphasia and Semantic Self-Monitoring

The lesions of people with Wernicke's aphasia typically occur in the posterior section of the superior temporal gyrus, in the person's language-dominant hemisphere. This part of the brain houses Wernicke's area and the primary auditory cortex (also known as Brodmann areas 41 and 42). Unlike patients with Broca's aphasia, those with Wernicke's aphasia have deficits that center on the comprehension of language. For this reason, Wernicke's is sometimes known as a *receptive* aphasia. Yet, unlike an ordinary person who, hearing an unfamiliar language, instantly recognizes that he does not understand, people with significant Wernicke's aphasia often seem not to realize that they are not understanding much of what is said to them.

To simply call Wernicke's a receptive aphasia, however, threatens to mask the fact that such patients have significant language-production problems as well, albeit of a different sort than those with Broca's or conduction aphasia. While the mere act of speaking comes easily to Wernicke's patients—making theirs a *fluent* aphasia—their speech is often unintelligible in ways unlike that of patients with aphasias that pertain primarily to language production. This is because the difficulties that Wernicke's patients have with language comprehension apparently extend to the comprehension

¹⁷ In a fascinating unpublished report, Arciniegas (unpublished) describes a patient with schizophrenia whose longstanding AVHs seemed to be eliminated after a stroke that left him with global aphasia. (Patients with global aphasia show language difficulties characteristic of *both* Broca's and Wernicke's aphasia). The patient's widespread lesions were to primary auditory cortex, the lateral temporal cortex, and perisylvian areas. This patient's lesions and aphasia do not provide a crucial test of the theories in question. Both would predict that lesions to primary auditory cortex will lead to a decrease in AVHs.

of their own speech as well. Their speech tends to be littered with misused words, non-words, neologisms, and paragrammatic language that render their utterances inscrutable, or simply meaningless (LaPointe 2005). There is a sense in which they “fluently” produce speech, without being aware of what they are in fact saying. Indeed, their manner often suggests that they have very clearly in mind what they wish to say, and simply fail to realize that the words they utter do not convey that information.

To give an example, drawn from many available on video through the AphasiaBank.org database (MacWhinney et al. 2011), a man with a diagnosis of post-stroke Wernicke's aphasia is asked by his therapist to say what he remembers about having his stroke. He carefully considers the question and responds:

You know to say that I'm only like two hours that were really from the first two back of...opens this...years back, January and March. Those are the flowers, this is the hardest it's been here. And this is the hardest way to strike, and I say, today I say, well, it's just open, and I should be talking, but I need somebody now, and open, what I've too, because I say, *oh my god*, this is awful. So much in front of me that if I walked away from it and touching everything and saying ok tomorrow's. I can open tomorrow, I can open things that I want to say more things to. Because these are so short open to me now, so much short, that need to want more. (MacWhinney et al. 2011).

A person who watched this interchange, but knew no English, would find little amiss. The patient appears focused, composed, purposeful, and in control of his faculties. His speech is marked by expressive pauses and emphases, such as when he says “*oh my god*” in a hushed and surprised tone, to convey astonishment. And yet his speech, though easy to decipher into individual words, conveys little in the way of ordinary semantic meaning. One has the impression of a person who knows exactly what he wishes to say, and who believes himself to be saying it, but who is in fact saying something else, or nothing at all.

As noted, the speech of people with Wernicke's aphasia is also often marked by the use of non-words or badly garbled words, interspersed among words pronounced with ordinary clarity. For instance, in another example from MacWhinney et al. (2011), when a patient is asked to describe what he remembers first happening after his stroke, he replies (non-words are in *italics*):

Good, generally good. But *wolking* this the first when my *begvole*...forget it. Because I had no... my foot was *fizzles* and stuff...I couldn't talk.

While this patient shows some difficulty finding the words he wishes to say, he does not evince any awareness of the particular moments where his speech deviates from ordinary English words. He does not, for instance, try to correct himself, or repeat what he has said in another way.

We can see the peculiarities of the speech in such examples as deriving from a failure of self-monitoring: the speaker aims to convey certain information, but fails to accurately monitor whether his speech in fact conveys that information. Yet this failure of self-monitoring is not usefully modelled as a failure to appropriately use motor commands to predict or anticipate sensory input, *à la* the forward model or filtering

architectures. The problem is not one of perceived speech sounds failing to match predicted speech sounds, or of failing to attenuate or filter auditory-phonological representations. The problem has to do with the *meaning* of the words not matching what the person intends to convey. This way of describing the situation presumes that such a person has clear intentions about what he wishes to express, even if he cannot find the proper words with which to express them. More conservatively, we can say that the language production and comprehension systems—both evidently needed for generating fluent and coherent speech—are not working together in an appropriately coordinated fashion. And the patient's nervous system fails to monitor and correct this lack of coordination.

As evidenced most saliently by the language-production abilities of people with damage to Broca's area, regions of the inferior frontal gyrus are relied upon simply to generate speech. And Wernicke's area, and the superior temporal gyrus more generally, has the more ambiguous role of ensuring that the speech one generates is semantically appropriate. In a sense, Wernicke's area can be thought of as monitoring the products of Broca's area, and shaping them as needed. However, it would be a mistake to conceive of Wernicke's area as a kind of inner ear, listening in on the productions of Broca's area. For the interaction between the two areas appears continuous and entirely essential for the generation of anything like ordinary coherent speech. And indeed, electrical readings taken directly from the respective areas in patients awaiting surgery for epilepsy show that the two areas form a "bi-directional" feed-forward and feed-back system, where artificial activation in one area leads to activation in the other, and vice versa (Matsumoto et al. 2004, p. 2316). Thus, Wernicke's area is not passively "listening" for errors and triggering an error signal only when one is detected. It is constitutively involved in language production, continuously shaping the output from Broca's area and feeding it back for further processing.

What relevance does this have to AVHs and thought insertion? While I do not suggest that people suffering AVHs or inserted thoughts have the equivalent of Wernicke's aphasia, some useful parallels can be drawn. Wernicke's aphasia shows us how a person can generate speech that is in one sense purposeful (insofar as it is not like the speech that characterizes Tourette's syndrome), but that does not match anything she intended to say. Further, it shows that we can be quite unaware of this mismatch—unaware that we meant to say one thing but in fact said another. The same could be true in the case of AVHs and inserted thoughts if both result, as many have proposed, from the generation of inner speech. The person suffering AVHs and inserted thoughts purposefully generates speech—either in the sense that motor commands lead to the relevant activation in Broca's area, or intentionally—and yet the resulting inner speech episode is not recognized as having been one's own doing, partly because it does not convey any information the person meant to express. That is, the actual meaning of the inner speech utterance may not match anything the person intended to say. And it may not match due to abnormalities in communication between Broca's and Wernicke's areas.

Interestingly, there is already considerable evidence that Wernicke's area, and the superior temporal gyrus more generally, is abnormally activated in people with AVHs. Hoffman et al. (2005) and Hoffman et al. (2013) found that transcranial magnetic stimulation of Wernicke's area helps curtail hallucinations in patients with persistent AVHs, while Homan et al. (2013) found abnormally increased cerebral blood flow in

superior temporal gyrus in patients with medication-resistant AVHs (see also Simons et al. (2010)). And, investigating the density of white matter tracts connecting brain regions, Hubl et al. (2004) found that patients with AVHs had significantly decreased density in pathways leading from Broca's area to Wernicke's and other nearby speech comprehension areas, suggesting abnormal patterns of coactivation between these areas. They argue that "the aberrant connections may lead to abnormal activation in regions that normally process external acoustical and language stimuli," which, "accounts for these patients' inability to distinguish self-generated thoughts from external stimulation" (p. 667). In a similar vein, de Weijer et al. (2011) found degraded integrity in the arcuate fasciculi (the main connection between language production and comprehension areas) in schizophrenia patients with chronic hallucinations. Other neuroimaging evidence indicates that people with schizophrenia have increased activity in left superior temporal areas during inner speech compared to controls (Simons et al. 2010), which again suggests a general propensity for abnormally monitoring and decoding the semantic aspects of speech. Similarly, conducting a meta-analysis of imaging data, Kompus et al. (2011) found that the superior temporal gyrus was unusually active among patients with schizophrenia during periods when they reported experiencing voices, yet abnormally inactive during ordinary acts of auditory perception.

It is possible interpret these results as support for something like the filtering or attenuation hypothesis discussed above—indeed, this appears to be how such studies are often perceived. For instance, Moseley et al. (2013) see the decreased connectivity between language production and perception areas as indicating a failure, during speech, to "dampen activity in auditory cortex and label it as self-generated" (p. 2797). And in the passage from Hubl et al. (2004) cited above, the authors hold that the miscommunication between language areas results in an inability to distinguish "self-generated thoughts from external stimulation." The suggestion again is that inner speech is not appropriately dampened or attenuated.

But the nature of Wernicke's aphasia should lead us to consider other possibilities. For it shows that the sort of "monitoring" accomplished by Wernicke's and related speech-perception areas during ordinary speech is not simply a matter of determining the causal source of a speech representation, based on features such as its sensory vivacity. These "reception" areas are also crucial for the *generation* of semantically coherent strings of speech. And it may be the semantic mismatch of one's speech with one's speech intentions that, in people with schizophrenia, at least partly contributes to its seeming to derive from an alien agency. For the patient would be in the position of being aware of various utterances represented in inner speech, without having had any intention to generate such utterances, and without their cohering with anything one might ordinarily say to oneself. A quick abductive inference might lead one to conclude, or at least raise one's suspicions that, the utterance is the product of someone else's agency.

To better-appreciate the semantic self-monitoring proposal, it may help to note the following contrasts to the motor-control self-monitoring views (such as the filter model) discussed above. On a motor-control self-monitoring view, the linguistic content of an AVH (i.e., what the voice is *saying* or *expressing*) is determined by a motor command to generate an utterance with that content. That motor command may or may not itself be caused by an intention to generate such an utterance (recall the contrast between an intention and a motor command, in Section 2). What accounts for

the sense that the resulting inner speech episode was caused by someone else is a malfunction in the ordinary processes by which self-created sensory stimuli are attenuated or filtered. Due to the malfunction, the inner speech episode becomes unusually “vivid” or salient. However, such errors in sensory attenuation will not themselves affect the linguistic content of the AVH; they will not affect what it is that the voice is *saying*. By contrast, on the semantic self-monitoring model, the linguistic content of an AVH will not be matched or straightforwardly determined by the content of an intention or a motor command. For while a motor command or intention to generate inner speech will *initiate* an episode that becomes an AVH, the transformation of ordinary inner speech into an AVH will be due to irregularities in how the semantic aspects of the utterance are developed, through ongoing interaction between speech production and comprehension areas. The result, when an AVH occurs, is an inner speech utterance that is out of keeping—in terms of its linguistic and semantic content—with any motor command or intention that the subject had. It is this kind of semantic anomaly that, on the semantic self-monitoring account, generates a sense that the AVHs was the result of some other agency. Having noted this contrast, however, it remains possible that *both* sorts of self-monitoring deficits might co-occur. These need not be viewed as competing hypotheses.

4.3 Semantic Self-Monitoring and Formal Thought Disorder

Further support for the semantic self-monitoring theory is available when we consider that people with schizophrenia often show comparable abnormalities in their ordinary overt speech as well. It is on the basis of exhibiting speech that is disjointed, inscrutable, and semantically impoverished that patients are diagnosed with *formal thought disorder* (FTD).¹⁸ Patients with FTD show much the same kind of disorganized speech that characterizes Wernicke’s aphasia. And, in both cases, speakers often show little awareness of their own speech abnormalities. So there is room to see both AVHs and formal thought disorder as resulting from improper connectivity between, or functioning in, language production and perception areas—specifically insofar as language “perception” areas are constitutively involved in language production as well.

It is a peculiarity of the existing literature on self-monitoring and AVHs that so little of it draws any connection between AVHs and the abnormal *overt* speech that often characterizes schizophrenia. For the two can easily be seen as resulting from a single malfunctioning language system (Hoffman 1986; Strik and Dierks 2008). Moreover, patients suffering severe forms of each symptom have shown similar neural volume decreases in the left superior temporal gyrus (Sun et al. 2009). Using whole-brain MRI with optimized voxel-based morphometry, Sans-Sansa et al. (2013) found FTD to be correlated with volume reductions in both Broca’s and Wernicke’s areas, mirroring some of the neural density results discussed above with respect to AVHs. and Sommer et al. (2010) found that non-pathological populations reporting AVHs showed positive FTD symptoms at comparable rates to schizophrenia patients with AVHs. Insofar as the two symptoms similarly co-occur in each population, this is suggestive of a shared, or strongly related neurobiology.

¹⁸ McKenna and Oh (2005) provide an excellent overview of the relation between formal thought disorder and language deficits in schizophrenia

The failure of most self-monitoring accounts to draw any etiological connection between AVHs and formal thought disorder may be due in part to the fact that factor analyses of symptom clusters tend to place AVH and FTD in separate clusters (Andreasen et al. 1995). Yet, as noted by Sommer et al. (2010), this shows at most that the symptoms do not tend to co-occur in the same patients *at the same time*. It may be that each symptom reflects a different stage of a progressive disease, with FTD being more severe. Also possible is that FTD is more responsive to medication than AVHs, which would create the illusion of a lack of correlation. And, in any event, a more recent dimensional analysis of symptom clusters, surveying over 500 patients, places both AVH and FTD within a single cluster labelled “reality distortion” (Demjaha et al. 2009). Further studies that look at brain activation and morphology with respect to the two disorders are essential to properly untangling their relation. The data to date certainly warrant further investigation into their connection.

In sum, there are plausible ways of understanding what self-monitoring of inner speech might involve that do not invoke the idea of sensory attenuation or filtering. There is a kind of semantic self-monitoring—involving interaction between distinct language regions—that is crucial for the generation of coherent speech that matches one's intentions, which may be impaired in patients with AVHs. The current focus of self-monitoring theories on models deriving from the motor control and perception literatures, combined with a failure to appreciate the role of the superior temporal gyrus in language *production* (as opposed to simply perception), has unfortunately obscured this alternative possibility for understanding the self-monitoring errors characteristic of AVHs and inserted thoughts.

5 Conclusion

I have proposed two ways of understanding the kinds of self-monitoring deficits in schizophrenia that may lead to reports of AVHs and inserted thoughts. This first is the filtering hypothesis. Like other approaches that posit deficits in attenuating or dampening self-created sensory signals, the filtering hypothesis holds that the inner speech of patients suffering AVHs and inserted thought becomes abnormally salient, due to a lack of the attenuation normally applied to such self-created signals. It is important, I argued, to understand how such a model can be applied to AVHs and inserted thoughts without bringing with it the questionable apparatus of a comparator mechanism.

An alternative to this approach sees Wernicke's aphasia as a telling analogy to the AVHs and inserted thoughts experienced by many with schizophrenia. The process by which we generate meaningful speech involves intricate and ongoing interaction between neural regions ordinarily associated with language production and perception, respectively. Deficits in the interaction between these areas can lead to speech that is at once self-generated yet semantically anomalous with respect to one's larger goals and intentions. The failure of such coordination, together with a failure to notice or correct this coordination, can be considered an alternative kind of monitoring deficit—one that may lend one's own inner speech an alien quality. A virtue of this approach is that it allows for etiological and explanatory links between the abnormal inner *and* overt speech observed in schizophrenia.

It is possible that both kinds of self-monitoring deficit contribute concurrently to the experience of AVHs and thought insertion. Nevertheless, while the neurocognitive hypotheses proposed here may shed some light on particular symptoms, they leave very many aspects of schizophrenia untouched, including its progression, its full range of symptoms (including negative symptoms), and its overall etiology. Indeed, many leading theories of schizophrenia sidestep cognitive-architectural forms of explanation altogether, focusing instead on irregularities in the function of neurotransmitters such as dopamine (Howes and Kapur 2009), or on genetic markers associated with the disorder (Arnedo et al. 2014). Nor is there consensus over whether the many symptoms lumped together under the heading of schizophrenia are in fact manifestations of a single underlying disorder, or several.

In approaching the issue from a cognitive-architectural perspective, we can hope to establish explanatory links between outward symptoms and lower-level phenomena, such as brain chemistry and DNA. Such links may help explain features of the symptoms themselves, such as their intermittency, by showing how neural abnormalities modulate a symptom's manifestation. At the same time, our higher-level understanding of the symptoms of schizophrenia can helpfully drive inquiry at lower levels, as in the case where the decreased density of neural networks linking language production and language perception areas was discovered through testing the hypothesis that AVHs result from a failure to properly monitor inner speech. Through such continued reciprocal investigations we may eventually arrive at an understanding of the symptoms of schizophrenia that allows for more effective intervention, while gaining a clearer view of the fault lines among ordinary thought, speech, and action.

Acknowledgments I am grateful to two anonymous reviewers for this journal, to Wayne Wu, and to Aimee Dietz, for their helpful comments, criticisms, and suggestions. This research was supported by a grant from the Taft Research Center at the University of Cincinnati.

References

- Alderson-Day, B., S. McCarthy-Jones, S. Bedford, H. Collins, H. Dunne, C. Rooke, and C. Fernyhough. 2014. Shot through with voices: Dissociation mediates the relationship between varieties of inner speech and auditory hallucination proneness. *Consciousness and Cognition* 27: 288–296. doi:10.1016/j.concog.2014.05.010.
- Allen, P., G. Modinos, D. Hubl, G. Shields, A. Cachia, R. Jardri, and R. Hoffman. 2012. Neuroimaging auditory hallucinations in schizophrenia: From neuroanatomy to neurochemistry and beyond. *Schizophrenia Bulletin* 38(4): 695–703. doi:10.1093/schbul/sbs066.
- Andreasen, N. C., S. Amdt, R. Alliger, D. Miller, and M. Flaum. 1995. Symptoms of schizophrenia. Methods, meanings, and mechanisms. *Arch Gen Psychiatry* 52(5): 341–351.
- Andreasen, N.C., and M. Falum. 1991. Schizophrenia: The characteristic symptoms. *Schizophrenia Bulletin* 17: 27–50.
- Arnedo, J., D.M. Svrakic, C. Del Val, R. Romero-Zaliz, H. Hernandez-Cuervo, A.H. Fanous, and I. Zwir. 2014. Uncovering the hidden risk architecture of the schizophrenias: Confirmation in three independent genome-wide association studies. *The American Journal of Psychiatry*. doi:10.1176/appi.ajp.2014.14040435.
- Aziz-Zadeh, L., L. Cattaneo, M. Rochat, and G. Rizzolatti. 2005. Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. *Journal of Cognitive Neuroscience* 17(6): 928–938. doi:10.1162/0898929054021157.

- Blakemore, S.J., D.A. Oakley, and C.D. Frith. 2003. Delusions of alien control in the normal brain. *Neuropsychologia* 41: 1058–1067.
- Blakemore, S.J., D.M. Wolpert, and C.D. Frith. 1998. Central cancellation of self-produced tickle sensation. *Nature Neuroscience* 1: 635–640.
- Blakemore, S.J., D.M. Wolpert, and C.D. Frith. 2002. Abnormalities in the awareness of action. *TRENDS in Cognitive Science* 6: 237–241.
- Christoffels, I.K., V. van de Ven, L.J. Waldorp, E. Formisano, and N.O. Schiller. 2011. The sensory consequences of speaking: Parametric neural cancellation during speech in auditory cortex. *PloS One* 6(5): e18307. doi:10.1371/journal.pone.0018307.
- de Weijer, A.D., R.C.W. Mandl, K.M.J. Diederens, S.F.W. Neggers, R.S. Kahn, H.E.H. Pol, and I.E.C. Sommer. 2011. Microstructural alterations of the arcuate fasciculus in schizophrenia patients with frequent auditory verbal hallucinations. *Schizophrenia Research* 130(1): 68–77. doi:10.1016/j.schres.2011.05.010.
- Demjaha, A., K. Morgan, C. Morgan, S. Landau, K. Dean, A. Reichenberg, and P. Dazzan. 2009. Combining dimensional and categorical representation of psychosis: The way forward for DSM-V and ICD-11? *Psychological Medicine* 39(12): 1943–1955. doi:10.1017/S0033291709990651.
- Ford, J.M., and D.H. Mathalon. 2005. Corollary discharge dysfunction in schizophrenia: Can it explain auditory hallucinations. *International Journal of Psychophysiology* 58(2–3): 179–189.
- Ford, J.M., D.H. Mathalon, T. Heinks, S. Kalba, W.O. Faustman, and W.T. Roth. 2001. Neurophysiological evidence of corollary discharge dysfunction in schizophrenia. *American Journal of Psychiatry* 158(12): 2069–2071. doi:10.1176/appi.ajp.158.12.2069.
- Frith, C. 1992. *The cognitive neuropsychology of schizophrenia*. Hove: Lawrence Erlbaum.
- Frith, C.D. 2005. The neural basis of hallucinations and delusions. *Comptes Rendus Biologies* 328: 169–175.
- Frith, C. 2012. Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition* 21: 52–54.
- Frith, C.D., and D.J. Done. 1989. Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine* 19(02): 359–363. doi:10.1017/S003329170001240X.
- Frith, C.D., S.J. Blakemore, and D.M. Wolpert. 2000. Abnormalities in the awareness and control of action. *Philosophical Transactions of the Royal Society, B: Biological Sciences* 355(1404): 1771–1788.
- Geva, S., P.S. Jones, J.T. Crinion, C.J. Price, J.-C. Baron and E.A. Warburton. 2011a. The neural correlates of inner speech defined by voxel-based lesion–symptom mapping 134.
- Geva, S., P.S. Jones, J.T. Crinion, C.J. Price, J.C. Baron, and E.A. Warburton. 2011b. The neural correlates of inner speech defined by voxel-based lesion-symptom mapping. *Brain* 134(Pt 10): 3071–3082. doi:10.1093/brain/awr232.
- Graham, G., and G.L. Stephens. 2000. *When self-consciousness breaks*. Cambridge: MIT Press.
- Grueschow, M., R. Polania, T.A. Hare, and C.C. Ruff. 2015. Automatic versus choice-dependent value representations in the human brain. *Neuron*. doi:10.1016/j.neuron.2014.12.054.
- Grush, R. 2004. The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences* 27: 377–442.
- Heinks-Maldonado, T. H., S. S. Nagarajan, and J. F. Houde. 2006. Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport* 17(13): 1375–1379. doi:10.1097/01.wnr.0000233102.43526.e9
- Hoffman, R. 1986. Verbal hallucinations and language production processes in schizophrenia. *Behavioral and Brain Sciences* 9(03): 503–517. doi:10.1017/S0140525X00046781.
- Hoffman, R., R. Gueorguieva, K.A. Hawkins, M. Varanko, N.N. Boutros, Y.-t. Wu, and J.H. Krystal. 2005. Temporoparietal transcranial magnetic stimulation for auditory hallucinations: safety, efficacy and moderators in a fifty patient sample. *Biological Psychiatry* 58(2): 97–104. doi:10.1016/j.biopsych.2005.03.041.
- Hoffman, R., M. Varanko, J. Gilmore, and A.L. Mishara. 2008. Experiential features used by patients with schizophrenia to differentiate ‘voices’ from ordinary verbal thought. *Psychological Medicine* 38: 1167–1176.
- Hoffman, R., T. Fernandez, B. Pittman, and M. Hampson. 2011. Elevated functional connectivity along a corticostriatal loop and the mechanism of auditory/verbal hallucinations in patients with schizophrenia. *Biological Psychiatry* 69(5): 407–414. doi:10.1016/j.biopsych.2010.09.050.
- Hoffman, R., K. Wu, B. Pittman, J.D. Cahill, K.A. Hawkins, T. Fernandez, and J. Hannestad. 2013. Transcranial magnetic stimulation of Wernicke’s and right homologous sites to curtail “voices”: A randomized trial. *Biological Psychiatry* 73(10): 1008–1014. doi:10.1016/j.biopsych.2013.01.016.
- Hohwy, J. 2013. *The predictive mind*. Oxford: Oxford University Press.

- Homan, P., J. Kindler, M. Hauf, S. Walther, D. Hubl, and T. Dierks. 2013. Repeated measurements of cerebral blood flow in the left superior temporal gyrus reveal tonic hyperactivity in patients with auditory verbal hallucinations: a possible trait marker. *Frontiers in Human Neuroscience* 7: 304. doi:10.3389/fnhum.2013.00304.
- Howes, O.D., and S. Kapur. 2009. The dopamine hypothesis of schizophrenia: version III—the final common pathway. *Schizophrenia Bulletin* 35(3): 549–562. doi:10.1093/schbul/sbp006.
- Hubl, D., T. Koenig, W. Strik, et al. 2004. Pathways that make voices: White matter changes in auditory hallucinations. *Archives of General Psychiatry* 61(7): 658–668. doi:10.1001/archpsyc.61.7.658.
- Hurlburt, R.T., and E. Schwitzgebel. 2007. *Describing inner experience? proponent meets skeptic*. Cambridge: MIT Press.
- Jones, S.R., and C. Fernyhough. 2007a. Neural correlates of inner speech and auditory verbal hallucinations: A critical review and theoretical integration. *Clinical Psychology Review* 27: 140–154.
- Jones, S.R., and C. Fernyhough. 2007b. Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition* 16(2): 391–399. doi:10.1016/j.concog.2005.12.003.
- Junginger, J., and C.L. Frame. 1985. Self-report of the frequency and phenomenology of verbal hallucinations. *The Journal of Nervous and Mental Disease* 173(3): 149–155.
- Kompus, K., R. Westerhausen, and K. Hugdahl. 2011. The “paradoxical” engagement of the primary auditory cortex in patients with auditory verbal hallucinations: a meta-analysis of functional neuroimaging studies. *Neuropsychologia* 49(12): 3361–3369. doi:10.1016/j.neuropsychologia.2011.08.010.
- Kühn, S., and J. Gallinat. 2010. Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. *Schizophrenia Bulletin*. doi:10.1093/schbul/sbq152.
- Langland-Hassan, P. 2008. Fractured phenomenologies: Thought insertion, inner speech, and the puzzle of extraneity. *Mind & Language* 23(4): 369–401.
- LaPointe, L.L. 2005. *Aphasia and related neurogenic language disorders*, 3rd ed. New York: Theime Medical Publishers.
- Laroi, F., I.E. Sommer, J.D. Blom, C. Fernyhough, D.H. Ffytche, K. Hugdahl, and F. Waters. 2012. The characteristic features of auditory verbal hallucinations in clinical and nonclinical groups: state-of-the-art overview and future directions. *Schizophrenia Bulletin* 38(4): 724–733. doi:10.1093/schbul/sbs061.
- Lindner, A., P. Thier, T.T.J. Kircher, T. Haarmeier, and D.T. Leube. 2005. Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Current Biology* 15(12): 1119–1124. doi:10.1016/j.cub.2005.05.049.
- MacWhinney, B., D. Fromm, M. Forbes, and A. Holland. 2011. AphasiaBank: Methods for studying discourse. *Aphasiology* 25(11): 1286–1307. doi:10.1080/02687038.2011.589893.
- Matsumoto, R., D.R. Nair, E. LaPresto, I. Najm, W. Bingaman, H. Shibusaki, and H.O. Lüders. 2004. Functional connectivity in the human language system: A cortico-cortical evoked potential study 127.
- McCarthy-Jones, S., and C. Fernyhough. 2011. The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults. *Consciousness and Cognition* 20(4): 1586–1593. doi:10.1016/j.concog.2011.08.005.
- McGuire, P.K., G.M. Shah, and R.M. Murray. 1993. Increased blood flow in Broca’s area during auditory hallucinations in schizophrenia. *Lancet* 342: 703–706.
- McGuire, P.K., D.A. Silbersweig, R.M. Murray, A.S. David, R.S.J. Frackowiak, and C.D. Frith. 1996. Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine* 26(01): 29–38. doi:10.1017/S0033291700033699.
- McKenna, P., and T. Oh. 2005. *Schizophrenic speech: Making sense of bathroosts and ponds that fall in doorways*. Cambridge: Cambridge University Press.
- Mellor, C.H. 1970. First rank symptoms of schizophrenia. *British Journal of Psychology* 117: 15–23.
- Miall, R.C., D.J. Weir, D.M. Wolpert, and R.C. Stein. 1993. Is the cerebellum a Smith Predictor? *Journal of Motor Behavior* 25: 203–216.
- Moseley, P., and S. Wilkinson. 2014. Inner speech is not so simple: a commentary on Cho and Wu (2013). *Frontiers in Psychiatry* 5. doi: 10.3389/fpsy.2014.00042.
- Moseley, P., C. Fernyhough, and A. Ellison. 2013. Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neuroscience and Biobehavioral Reviews* 37(10 Pt 2): 2794–2805. doi:10.1016/j.neubiorev.2013.10.001.
- Nayani, T.H., and A. David. 1996. The auditory hallucination: A phenomenological survey. *Psychological Medicine* 26: 177–189.
- Oppenheim, G.M. 2013. Inner speech as a forward model? *Behavioral and Brain Sciences* 36(04): 369–370. doi:10.1017/S0140525X12002798.
- Penfield, W., and P. Perot. 1963. The brain’s record of auditory and visual experience. *Brain* 86(4): 595.

- Pickering, M.J., and S. Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36(04): 329–347. doi:10.1017/S0140525X12001495.
- Raij, T.T., and T.J.J. Riekkki. 2012. Poor supplementary motor area activation differentiates auditory verbal hallucination from imagining the hallucination. *NeuroImage Clinical* 1(1): 75–80. doi:10.1016/j.nicl.2012.09.007.
- Sans-Sansa, B., P.J. McKenna, E.J. Canales-Rodriguez, J. Ortiz-Gil, L. Lopez-Araquistain, S. Sarro, and E. Pomarol-Clotet. 2013. Association of formal thought disorder in schizophrenia with structural brain abnormalities in language-related cortical regions. *Schizophrenia Research* 146(1–3): 308–313. doi:10.1016/j.schres.2013.02.032.
- Sartorius, N., R. Shapiro, and A. Jablonsky. 1974. The international pilot study of schizophrenia. *Schizophrenia Bulletin* 1: 21–35.
- Sasin, E., M. Nieuwenstein, and A. Johnson. 2015. The role of depth of encoding in attentional capture. *Psychonomic Bulletin and Review*. doi:10.3758/s13423-015-0807-6.
- Schneider, K. 1959. *Clinical psychopathology*. New York: Grune and Stratton.
- Schwitzgebel, E. 2008. The unreliability of naive introspection. *The Philosophical Review* 117(2): 245–273.
- Shergill, S.S., G. Samson, P.M., Bays, C.D., Frith, and D.M. Wolpert. 2005. Evidence for sensory prediction deficits in schizophrenia. *Am J Psychiatry* 162(12): 2384–2386. doi:10.1176/appi.ajp.162.12.2384
- Shergill, S.S., T.P. White, D.W. Joyce, P.M. Bays, D.M. Wolpert, and C.D. Frith. 2014. Functional magnetic resonance imaging of impaired sensory prediction in schizophrenia. *JAMA Psychiatry* 71(1): 28–35. doi:10.1001/jamapsychiatry.2013.2974.
- Simons, C.J., D.K. Tracy, K.K. Sanghera, O. O'Daly, J. Gilleen, M.D. Dominguez, and S.S. Shergill. 2010. Functional magnetic resonance imaging of inner speech in schizophrenia. *Biological Psychiatry* 67(3): 232–237. doi:10.1016/j.biopsych.2009.09.007.
- Sommer, I.E., and K.M. Diederer. 2009. Language production in the non-dominant hemisphere as a potential source of auditory verbal hallucinations. *Brain* 132(10): e124–e124. doi:10.1093/brain/awp040.
- Sommer, I.E., A.M. Derwort, K. Daalman, A.D. de Weijer, P.F. Liddle, and M.P. Boks. 2010. Formal thought disorder in non-clinical individuals with auditory verbal hallucinations. *Schizophrenia Research* 118(1–3): 140–145. doi:10.1016/j.schres.2010.01.024.
- Sperry, R.W. 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. *The Journal of Comparative and Physiological Psychology* 43: 482–489.
- Stephane, M., S. Barton, and N.N. Boutros. 2001. Auditory verbal hallucinations and dysfunction of the neural substrates of speech. *Schizophrenia Research* 50(1–2): 61–78. doi:10.1016/S0920-9964(00)00150-X.
- Strik, W., and T. Dierks. 2008. Neurophysiological mechanisms of psychotic symptoms. *European Archives of Psychiatry and Clinical Neuroscience* 258(Suppl 5): 66–70. doi:10.1007/s00406-008-5016-0.
- Sun, J., J.J. Maller, L. Guo, and P.B. Fitzgerald. 2009. Superior temporal gyrus volume change in schizophrenia: a review on region of interest volumetric studies. *Brain Research Reviews* 61(1): 14–32. doi:10.1016/j.brainresrev.2009.03.004.
- Synofzik, M., G. Vosgerau, and A. Newen. 2008. Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition* 17(1): 219–239. doi:10.1016/j.concog.2007.03.010.
- Vicente, A. 2014. The comparator account on thought insertion: some open questions. *Phenomenology and the Cognitive Sciences* 13: 335–353.
- von Holst, E., and H. Mittelstadt. 1950/1973. The reafference principle (R. Martin, Trans.) *The Behavioral Physiology of Animals and Man. The Collected Papers of Erich von Holst*.
- Vosgerau, G., and A. Newen. 2007. Thoughts, motor actions, and the self. *Mind and Language* 22: 22–43.
- Weiskrantz, L., J. Elliot, and C. Darlington. 1971. Preliminary observations of tickling oneself. *Nature* 230: 598–599.
- Wilkinson, S. 2014. Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework. *Consciousness and Cognition* 30: 142–155. doi:10.1016/j.concog.2014.09.002.
- Wolpert, D.M., R.C. Miall, and M. Kawato. 1998. Internal models in the cerebellum. *TRENDS in Cognitive Science* 2: 338–347.
- Wu, W. 2012. Explaining schizophrenia: Auditory verbal hallucination and self-monitoring. *Mind and Language* 27(1): 86–107.
- Wu, W., and R. Cho. 2013. Mechanisms of auditory verbal hallucination in schizophrenia. *Frontiers in Schizophrenia*, 4.