

# Explaining Schizophrenia: Auditory Verbal Hallucination and Self-Monitoring

Wayne Wu

(Forthcoming, *Mind and Language*; please quote from published article)

**Abstract:** Do self-monitoring accounts, a dominant account of the positive symptoms of schizophrenia, explain auditory verbal hallucination? In this essay, I argue that the account fails to answer crucial questions any explanation of auditory verbal hallucination must address. Where the account provides a plausible answer, I make the case for an alternative explanation: auditory verbal hallucination is not the result of a failed control mechanism, namely failed self-monitoring, but, rather, of the persistent automaticity of auditory experience of a voice. My argument emphasizes the importance of careful examination of phenomenology as providing substantive constraints on causal models of the positive symptoms in schizophrenia.

---

I am very grateful to Ian Gold and a second, anonymous referee, for their very helpful suggestions, challenges and comments on this paper. Earlier versions were presented at the Center for the Philosophy of Science, University of Pittsburgh, at the Berlin School of Mind and Brain, Humboldt Universität zu Berlin, and at the Center for the Neural Basis of Cognition, Carnegie Mellon University and I am grateful for the many helpful comments I received. I am also grateful to Ray Cho, Casey O'Callaghan, Michael Pauen and Jason Rosenstock for advice that affected the content of the paper and to Ralph Hoffman, Stephen Moritz, and Massoud Stephane for answering questions about their work.

Address for Correspondence: Center for the Neural Basis of Cognition, 115 Mellon Institute, 4400 Fifth Avenue, Carnegie Melon University, Pittsburgh, PA 15213  
Email: waynewu@andrew.cmu.edu



## 1. Introduction

What would it be to explain schizophrenia, a psychological disorder that affects around 1% of the population? Among its salient features are the *positive symptoms*, including *delusions of control*, where the subject intentionally acts and yet claims that another is controlling them; *auditory verbal hallucinations* (AVH), where the subject hears a voice that is not actually present; and *thought insertion*, where the subject experiences a thought and paradoxically claims that it belongs to another. An explanation of schizophrenia would be in part an explanation of these experiences. In this essay, I focus on AVH and assess the adequacy of self-monitoring accounts, a dominant family of explanations of the positive symptoms. I argue that these models fail to answer or provide the wrong answer to questions any explanation of AVH must address. Furthermore, I make the case for an alternative account that focuses on automaticity as a central feature of AVH.

My approach will be to examine the phenomenology of AVH. After all, AVH involves odd experiences and to explain it, we need to systematically identify its features on pain of lacking a determinate explanatory target. We can then pose clearer questions and identify possible answers concerning the basis of AVH. Certainly, the issues regarding underlying mechanisms are ultimately empirical. My goal in this paper is not to systematically examine empirical studies but to identify substantive constraints on models of AVH via its experiential features. I will, however, make some general comments about empirical results below.

Here are three basic questions we can raise for any positive symptom  $F$ :

1. Identification: *What is symptom F?*
2. Phenomenological: *Why does F involve phenomenal character A?*
3. Causal: *Why does F occur in patients with schizophrenia?*

Self-monitoring accounts provide answers to these questions and are unified by positing a mechanism of self-monitoring that is defective in subjects with schizophrenia. I shall make a case for three points where these questions concern AVH: (1) self-monitoring accounts often rely on the wrong answer to the identification question; (2) they make incorrect predictions in respect of the phenomenological question and are not needed to answer it; and (3) while they offer a causal explanation of one feature of AVH, namely, involuntariness, a plausible alternative is available: AVH is automatic auditory experience of a voice. Given the significant explanatory shortcomings of self-monitoring accounts in respect of AVH, we should consider an alternative explanation.

I proceed as follows: In section 2, I articulate three questions that we must answer in explaining AVH. I argue, in section 3, that self-monitoring accounts misidentify AVH with inner speech. In section 4, I argue that they give the wrong answer to the phenomenological question, and I provide an answer that eschews self-monitoring in favor of the representational content of AVH. In section 5, I flesh out my account of AVH as automatic auditory experience of other voices by explicating the notion of automaticity. Finally, in section 6, I raise the possibility that the underlying mechanisms of the positive symptoms divide between agentic and non-agentic forms and accordingly are heterogeneous.

## 2. Three Questions about AVH

Upwards of 70% of subjects diagnosed with schizophrenia suffer from AVH where they 'hear voices'. But what is AVH? Let me begin with two clarifications. First, 'voice' (and 'speech') is unfortunately ambiguous. In this essay, by 'voice' (or 'speech'), I always refer to a type of *sound*, namely that physical entity produced in human utterances (scientists identify it with a sound wave). To hear a voice is to be in an auditory state that represents the presence of this sound, veridically or not. Psychologists and clinicians, however, sometimes mean 'voice' where it is *true* that there is a voice in the subject's head. As there is not literally such a sound in the head, 'voice' might here mean a type of thought as when one speaks of 'the voice of conscience'.<sup>1</sup> Related to this, a third meaning that preserves truth of talk of a voice in the head—one more often connected to use of '(inner) *speech*'—is the *experience* of a voice, an internal event that the subject fails to monitor. Finally, 'voice' also concerns the source of the putative sound heard (e.g. the speaker). In what follows, while self-monitoring accounts place emphasis on the last two readings, I am going to emphasize the auditory representation of voices in the first sense.

The first and third senses lead to a familiar but crucial distinction between the psychological *episode*, that is, the internal experience, and its representational *content*. In AVH, the internal episode is the hallucinatory experience, but its content, that which specifies what the experience represents, is the presence of a type of sound, a voice (philosophers speak of the representational *vehicle* versus

---

<sup>1</sup> Bleuer notes that 'patients are not always sure that they are actually hearing the voices or whether they are only compelled to think them. There are such "vivid thoughts" which are called voices by the patients' (quoted in Jones (2010, p.570)).

its *content*). So, when a subject is undergoing an episode of AVH, she is in a hallucinatory state that wrongly represents the presence of a voice. As this distinction is critical in what follows, consider another case: a symbol that flashes on a screen (episode) and what the symbol represents, its meaning. The symbol is different from its meaning (content) and we can ask separate questions concerning them: *How many times does the symbol occur on the screen? What is its meaning?*

Let us now answer the identification question:

- *Identification:* What is AVH?

My initial answer is literal: AVH is auditory verbal hallucination. Accordingly, in every instance of AVH, the relevant internal state is a type of auditory state, an experience that represents the presence of a speaking voice (this will be supported by the phenomenology considered below). Accordingly, AVH episodes so characterized count as hallucinations in that they represent objects that do not exist in the speaker's environment: there is no such voice at that time with the specific semantic content heard (clinicians speak of AVH as triggered in the absence of relevant external stimuli; my answer implies this given my use of 'voice'). We can broaden this account to encompass auditory hallucinations in general by loosening the restriction to voices (patients do not always hear just voices). I shall focus on the verbal case.

Given the distinction between episode and content, questions about AVH divide along properties specific to each. With regard to the episode, we can ask why the experience (event) has the duration that it exhibits, why it occurs when

it does, what triggers it, and why it is (seemingly) involuntary. With regard to the content, we can ask why the episode represents a voice and why it attributes to the voice the specific properties that it does, say audible properties, location, semantics and identity.

Are there additional features of AVH specific to or characteristic of schizophrenia? On the episode side, I want to mention one: AVH seems involuntary, not directly in one's control. This feature is widely attributed to AVH so as to become part of definitions of it (see Aleman and DeHaan's definition as discussed by David (2004) and Bentall et al. (1994), p. 52). In a report by Aggernaes et al. (1972), 84% of patients characterized their hallucinations as involuntary (number of subjects, n=45)<sup>2</sup> as did 100% of the patients in a study by Garrett and Silva (2003; n=32). It should be noted that this finding is not always so stark. Hoffman et al. (2007) found that about 28% of their patients claim to have no control of their AVH episodes and nearly 75% claim no or some control (n=44). On the other hand, Moritz and Larøi (2008) find 50% of their subjects (n=45) claimed predominant if not total control over their AVH episodes, specifically their content.<sup>3</sup> In any event, the sense of involuntariness is frequent enough to consider it a distinctive feature of AVH even if it is not always present.

---

<sup>2</sup> Four of these patients had non-auditory hallucinations; the others had auditory hallucinations. In the reports that follow, percentages are given in light of the number of subjects, n.

<sup>3</sup> Moritz, personal communication, cautions that their internet sample may not be representative of the general population of subjects with schizophrenia. As a nonclinical population, they may cope better with their AVHs (I am grateful to him for discussion). Their result is consistent with Nayani and David (1996): with 100 subjects (61% diagnosed with schizophrenia), 51% claim to exercise some control over their hallucinations; 38% claimed some ability to control the start of their voices, e.g., by asking it questions; 21% claimed to be able to sometimes stop their voices; however, 5% or less claimed to control content, speed and volume.

On the content side, the represented voices make utterances, typically commands and evaluations, that are generally negative in meaning, are directed to the patient (second-person) or speak about her in the third-person.<sup>4</sup> The voices are also experienced as having distinct audible properties (pitch, timbre, and intensity) and as having a specific gender, accent and identity (public figure, acquaintance, even the supernatural). In many cases, the voice is spatially located somewhere in the subject's environment though often, the voice seems to be emanating from a location *within the head*. Finally, there is a further feature that is often discussed, *alienness*: AVH has an 'alien' quality (Jones and Fernyhough, 2007, 391), involves experience of 'alien voices' (Aleman et al., 2003) or 'alienness/otherness' sensations (Hoffman, 2009, p.504) or itself is construed as 'alien' (Bentall et al., 1994, pp.52-3).

It is worth emphasizing the auditory aspect of AVH as there are reports of patients who experience voices that have no acoustical properties. For example, Moritz and Larøi (2008) report that 6% of their population afflicted with schizophrenia claim that the hallucinated voice is absolutely silent. Jones (2010) also notes a patient who reports 'soundless voices' (p. 570). These cases, however, do not count as *auditory* hallucinations and thus are not cases of AVH. It courts confusion to speak of *hearing* voices in this sense of 'voice' or one means something different by 'hearing' (as Moritz and Larøi also note, *op. cit.*, p. 104). The proper objects of auditory experience, what those experiences typically

---

<sup>4</sup> For a general survey of the phenomenology of AVH from which my description derives, see among others: Nayani and David (1996), Langdon et al. (2009), and Stephane et al. (2003).



represent, are sounds. As I focus on auditory hallucinations, I set aside these cases of soundless voices.<sup>5</sup>

Still, 'soundless voices' illustrate the diversity of phenomenology in the experiences of subjects with schizophrenia. For many of the features noted above, we can find cases of AVH that lack them. Thus, Larøi (2006) and Jones (2010) suggest that we should consider multiple mechanisms for AVH given the diversity of AVH experiences. I sympathize with this. Speculatively, such diversity might explain why imaging studies of AVH show striking variability. If scan groups differ in AVH phenomenology and hence in underlying mechanisms, the observed neural activity could be variable. For example, Copolov et al. (2003) using PET find activation of secondary auditory cortex and part of Wernicke's area but not primary auditory cortex or Broca's area during AVH. In contrast, van de Ven et al. (2005) find activation of primary auditory cortex in half their patients during AVH (n=6; fMRI) and McGuire et al. (1993), using PET, found hallucination related activity in Broca's area in 13 patients.<sup>6</sup> As we shall see, there is certainly diversity in AVH experience.

In the face of phenomenological diversity, explanation requires regimentation of target properties. Three features stand out: (1) the voices *sound* like voices, namely have audible properties; (2) AVH episodes are involuntary; and (3) they have an alien quality. To set a clear target, I suggest the following initial answer to the identification question: AVH is auditory verbal hallucination

---

<sup>5</sup> I am inclined to say that hallucinations are always perceptual (e.g. auditory) states and that with soundless voices, we have an intrusive *thought* phenomenon. But let that pass.

<sup>6</sup> We should investigate whether diversity of imaging results reflects diversity of phenomenology and underlying mechanisms. Van de Ven et al. (2005) suggest that primary auditory cortex activation might correlate with the vividness of AVH. Given consistently reported heard intensity of voices and variation along that dimension, this idea is in principle testable.

of a sound, a voice/speech, where in schizophrenia typically the episode is (or seems) involuntary and alien. This regimentation identifies features often found in AVH, and any complete account must explain them. Thus, we should address the following question:

- *Phenomenological: Why does AVH involve alienness?*

Finally, we are interested in the causal basis of AVH, specifically the mechanisms generating the audible properties of the experienced voice, its involuntariness and its alienness. We begin with the general causal question:

- *Causal: Why does AVH occur in patients with schizophrenia?*

Given our answer to the identification question, the causal question divides into three: (a) why does AVH represent a voice with the audible properties in question; (b) why is AVH involuntary; and (c) why does it involve alienness? The next sections assess how self-monitoring accounts fare on the questions I have identified. I begin with identification.

### **3. Self-Monitoring and AVH as defective Inner Speech: The Identification Question**

Irwin Feinberg (1978) was perhaps the first to deploy self-monitoring to explain the positive symptoms of schizophrenia, but the approach is widely endorsed

(see Cho and Carter (forthcoming) for a critical review). Thus, Allen, Aleman and McGuire note that:

[a] number of different theoretical frameworks have been proposed to account for the experience of AVH. Although these accounts have important differences they all propose that internally generated thoughts and images are mistaken for externally generated events ((2007), 409; see also Baker and Morrison (1998, p.p. 1199) for a similar statement).

The defining assumption of any self-monitoring account postulates a mechanism where internally generated events are identified as self-generated. Self-monitoring accounts deploy this mechanism, namely a defect in it, to explain AVH. A second assumption is that the positive symptoms amount to defective agency (see Seal et al., (2004), p. 49; Langdon et al., (2009) p. 655; Frank et al., (2000) p. 357; and Frith and Done (1988)). The failure of self-monitoring is a failure of a component of an action control system. I shall introduce self-monitoring models by presenting Christopher Frith's (1995) account, but let me emphasize that my target is the *family*.

Frith's account draws on a motor control model of Daniel Wolpert's (for an overview, see (Wolpert & Ghahramani 2000)). The model can be presented as follows:

Please see publication for figure

Figure 1: A schematic representation of self-monitoring in the motor control system. From C. Frith, S.-J. Blakemore, and D.M. Wolpert (2000) with permission.

The model notes that a copy of the motor command is sent to a prediction unit that has two functions. First, it compares predicted output with desired output. Should there be a mismatch, a correction is made, even before the motor action is generated. Second, the mechanism compares predicted output with actual output, and this is the locus of self-monitoring: where there is a match, the actual output is marked as self-generated; where there is not, trouble potentially ensues in patients with schizophrenia.

On many self-monitoring accounts, AVH results from a type of defective agency in that self-monitoring fails. But what kind of episode is involved? We need an answer to the identification question. Evans et al. note that '[p]sychological theories have reached a relative consensus in relating AVH to "inner speech"' (2000, p.137). Similarly, Jones and Fernyhough note that 'an assumption which has guided much thinking on AVHs [is] that they are a form of inner speech' (2007, 141) or 'verbal thought' (Allen et al, op. cit. p. 2007). These episodes are experiential in that there is something it is like to undergo them. We are then presented with the following explanation of AVH phenomenology. The subject generates inner speech but given a defect in self-monitoring, the speech is not identified as internally generated. Accordingly, the subject experiences her own inner speech as involuntary and alien.

Such self-monitoring accounts, focusing on the monitoring of inner episodes, endorse what I call the *strong (identification) thesis*, namely AVH is the subject's own inner speech misattributed to another. We can then think of an inner speech episode as an intermediate in the production of AVH experiences. Langdon et al. (2009) speak of inner speech as the 'raw materials' for AVH (p.

662) and Ditman and Kuperberg (2005), in a recent review of self-monitoring accounts, describe the posited misattribution as ‘the act of attributing an event to an incorrect source’ (p. 281). Shergill et al. (2001) write: ‘verbal hallucinations are derived from a lack of awareness of the subject’s normal inner speech’ (251).<sup>7</sup> The strong thesis implies but is not implied by the *moderate thesis*: some of the mechanisms involved in generating inner speech are involved in generating AVH.<sup>8</sup> I shall argue against the strong thesis.

The crucial logical point is that the moderate thesis can be true even if the strong thesis is false, e.g., if AVH is *not* inner speech. One way to think of the difference at issue is in a coarse description of two processes. We can represent inner speech, *B*, and AVH, *C*, as two types of internal experiences where on the strong thesis, the process is linear in the following way: there is an originating mechanism, *A*, that leads first to inner speech then, due to failure of self monitoring, yields AVH. Thus:  $A \rightarrow B \rightarrow C$ . Inner speech is the intermediate of AVH where both result from the same originating mechanism. On one version of the moderate thesis incompatible with the strong thesis, the originating mechanism (or in AVH, a proper part of it,  $A^*$ ) gives one of two distinct products where inner speech is not an intermediate of AVH. Thus:  $B \leftarrow A/A^* \rightarrow C$  (*B* and

---

<sup>7</sup> For similar characterizations, see Allen et al., (2008), p. 409, Johns et al., (2001) p. 705, Brebion et al., (2000) p. 119; Frank et al., (2000), p. 357; Keefe et al., (2002) p. 51 as well as papers by Frith, Blakemore and colleagues. What is at issue is the misattribution of an inner speech episode since it is internal events that self-monitoring mechanisms track. This implies the strong thesis.

<sup>8</sup> One reason for the link to inner speech is that subvocalization is often concomitant with both inner speech and AVH, and disruption of subvocalization has been reported to disrupt AVH (see Seal, Aleman, and McGuire (2004) and Ditman and Kuperberg (2005)). These observations are consistent with inner speech and AVH sharing a common mechanism that causes subvocalization (see Stephane et al. (2001)). On its own, it does not imply the strong rather than moderate thesis.

C are not simultaneously produced).<sup>9</sup> Here, AVH is not, nor is derived from, inner speech. Note that this model also predicts activation of similar brain areas during AVH and inner speech experiences.

Generically, we can think of inner speech as verbalized thought. There are three explications of this idea that I have found in the literature on AVH. The first, which I shall set aside, makes the identification of AVH as inner speech definitional: inner speech is *any* internal linguistic experience. Thus Leudar et al. (1997, p. 885-6) note that AVH counts as inner speech because it consists of '(i) word or sentences; which are (ii) heard as spoken; (iii) to the voice hearer; and (iv) which cannot be experienced directly by other people.' Since the identification with inner speech is supposed to be falsifiable, this characterization of inner speech cannot be what self-monitoring theorists have in mind.

Two narrower accounts are (a) inner speech as silent articulation of words and (b) Vygotskian conceptions of inner speech as internalized dialogue that serves self-regulation. A common instance of the first is mentally reciting sentences (see Jones and Fernyhough, 2007, sect. 2 for references). Here, inner speech as a silent form of articulation is experienced as producing speech. In Hurlburt (1990), his normal subjects often emphasize the productive aspects of 'inner speech' or that it is 'just like speaking aloud except no sound' (Hulbert and Schwitzgebel (2007, p.257)). The Vygotskian conception emphasizes the dialogic

---

<sup>9</sup> A\* is not strictly sufficient for C but might be when conjoined with auditory mechanisms. In principle, we should be able experimentally to tease the models apart. For example, if only a subset of the mechanisms sufficient for inner speech are deployed to generate AVH, we might find deficits in inner speech without concomitant defects in AVH. The current point, however, is to depict that on the moderate view, AVH occurs without inner speech as a precursor.

structure of inner speech, along with associated pragmatics (see Jones and Fernyhough, 2007 and Leudar et al., 2009 for discussion and references).<sup>10</sup>

I take no stand on which account is correct. My argument relies on the fact that inner speech episodes are internal experiences that we can recognize via reasonably consistent experiential properties even if we do not ourselves know the correct theory of their nature. One way then to show that AVH is not inner speech is to show that they differ in experiential properties. If there are sufficiently clear differences between the reported phenomenology, this is a *prima facie* problem for the strong thesis that takes inner speech as an intermediate to AVH. What we need, then, are careful analyses of phenomenology. In what follows, I report results based on phenomenological surveys of patients and normal subjects regarding AVH and inner speech.

I present the following challenge: to say that AVH is misattributed inner speech commits one to explaining how the differences between the phenomenal features of normal inner speech and those of AVH arise. We must explain this ‘transformation’ from the normal to the pathological. Yet it is implausible that a defect in self-monitoring explains it. In response, self-monitoring theorists can supplement their accounts with an additional mechanism to effect the transformation, but once this move is made, there is, I argue, an equally plausible alternative: AVH as auditory experience of another voice.

---

<sup>10</sup> A Vygotskian view might hold that inner speech can include cases of auditory imagination or recollection of dialogue, perhaps at an earlier stage in development. Inner speech loses auditory features when it matures. Jones and Fernyhough (2007) suggest that a return to the auditory form is the basis of AVH. This view of AVH is more along the lines of the auditory experience form of self-monitoring considered below and thus might not be committed to the strong thesis. If so, the current argument is directed at the inner articulation model.

There is a dearth of systematic study of the phenomenology of inner speech. Hulbert (1990) reports the phenomenology of inner speech in six normal subjects though Langdon et al. (2009) more recently address this issue with a greater number of normal and patient subjects (n=71, 29 with schizophrenia). Two of the authors in the latter report (Jones and Fernyhough) advocate inner speech models but conclude: 'if inner speech, conceived as the act of internal self-talk, is the raw material of all AVHs, then there should be similarities between the phenomenological characteristics of patients verbal thought and their AVHs. We found *no evidence* to support this prediction' (662; my italics).

First, it is not clear that inner speech involves *auditory* experience of a voice. In the study by Langdon et al., only a small number of their subjects reported that inner speech *sometimes* involved hearing a voice (12/49 of their normal control group; 11/29 of their schizophrenic group) (2009, p. 660). In contrast, AVH seems clearly to be the auditory experience as of a voice as sound, and this is reflected in subjective reports. In many studies where the phenomenology of AVH is probed, participants often report that 'hearing voices was very much like hearing other people' (100% in the sample in Leudar (1997, p.889), n= 28; cf. 93% of those questioned in Aggernaes (1972), n= 41; 100% of those sampled in Garrett and Silva (2003), n=32). Even where subjects claim a difference in 'reality' between hearing their voices and hearing other people, they largely affirm auditory features in their experience.<sup>11</sup> So, in AVH, we have either

---

<sup>11</sup> A referee draws my attention to possible conflicting data from Moritz and Larøi, 2008, who found that 38% of their subjects with schizophrenia indicate that the voices are not very real. This is, however, consistent with the experience being an auditory state in its representing certain audible properties like intensity (see next paragraph). Their subjects do not deny that the voices vary along audible intensity: 94% of their subjects indicate that the voices are at least whispering (16.2%), but most indicate that they are somewhat



the presence of auditory representation, one that picks out typical audible properties of voices, whereas in inner speech we do not, or we have in AVH a richer, more determinate auditory representation than in inner speech.

Let us pursue auditory determinacy in AVH versus inner speech. Consider audible intensity. It is not clear that loudness is represented in inner speech. MacKay, in a theoretical paper on inner speech, reports that subjects will report 'neutral' loudness and he concludes that loudness is not represented (MacKay 1992, p.128). Moritz and Larøi (2008) report that 80% of their normal subjects report absolute silence in their verbalized thinking. By implication, 20% of these subjects presumably take thoughts as having some auditory properties and the percentage increases among patients (39% report their thoughts somewhat audible; 61% as completely inaudible), so MacKay's claim is not generally true. Nevertheless, the contrast with AVH remains striking, for among Moritz and Larøi's population, 95% claim that their voices have an audible intensity if only at the level of whispering, so among a significant number of patients in their population (at least 55%), the contrast between thought and AVH is stark: AVH involves experienced intensity of sound, inner speech does not. The representation of intensity seems common in AVH. Moritz and Larøi (2008) report 74% of their subjects as hearing voices less loud than an external

---

audible, say as male or female voices (56.8%) or are as loud as external voices (21.6%). My point is that AVH experience is like hearing normal voices in its representing audible properties. Stephens and Graham (2003) report a study (Alpert & Silvers 1970) where AVH in alcoholics were compared with those in patients with schizophrenia and note that the authors 'found that alcoholics were more likely to mention specific 'sensory' features of the voice, such as its apparent volume or pitch, whereas 'hallucinations of schizophrenics have a more cognitive taint, appearing more like thoughts that have become audible' (26). Stephens and Graham cite other cases where 'voices' are said to have non-sensory features. Again, this just emphasizes that 'experiencing voices' can cover auditory and non-auditory cases.

voice (table 2, p. 101) and 16% as hearing whispering voices (n=45); Nayani and David (1996) report that 14% of their subjects described their voices as whispering, 13% as shouting, and 73% took their voices to be at normal conversational volumes (n=100).<sup>12</sup> What is clear is that AVH involves representation of audible intensity while inner speech often does not.

Voices in AVH have specific audible properties. Patients with schizophrenia often identify a specific gender with their voices. Nayani and David (1996) report male voices as more likely in their population: voices of a middle-aged male, n=34, a young adult male, n=24, and a young adult female, n=10; this was true in both male and female subjects (see table 3, op. cit., p. 181). Many of these voices (71%) are reported to have *accents* different from the patient's own. Moreover, the voices are represented as distinct in identity from the speaker and often are identified with known individuals, say family members, acquaintances or public figures (alive or dead; in Leudar et al., 1997, 46% were public figures, 23% acquaintances; 31% were supernatural characters; n=28). In contrast, for inner speech as reported by Hulbert's (1990) subjects, the voice of inner speech is reported as one's own voice. AVH involves representations of audible properties associated with another's voice.

There is also the spatial location of the voice. In Hulbert's (1990) subjects, inner speech was reported to be located in their head (e.g. see 'Helen', pp. 96ff). However, in AVH, there is often representation of the voice in external space.

---

<sup>12</sup> On Frith's model, self-monitoring leads to sensory attenuation of the predicted sensory consequences (see the discussion of self-tickling in Blakemore, Wolpert and Frith (2000)). It would then predict that in inner speech that is marked as self-generated, the 'intensity' of speech will be attenuated while in those where self-monitoring fails, the speech will not be attenuated and perhaps will appear of louder intensity. The model does not easily explain the experience in some AVH of the voice as whispering.

Hoffman et al. note that about 27% of schizophrenic patients they surveyed located the voice *exclusively* externally (2007) and he notes (personal communication) that about 50% of his surveyed patients at least sometimes locate the sounds externally.<sup>13</sup> Nayani and David (1996) report that 49% of their sample reported the voices as external, 38% as internal. Thus, AVH involves external localization of a voice while inner speech does not appear to.

The phenomenological data suggests that while both AVH and inner speech are experiences of language, the latter seems to be more often abstracted from an auditory format, namely without representation of audible properties. For example, inner speech may involve a more abstract linguistic representation, say of words but not of vocalized speech (cf. the Vygotskian account of condensed inner speech, Jones and Fernyhough, 2007, p. 147). Still, the line is not always sharp for some report inner thoughts with audible properties, specifically intensity. Nevertheless, in AVH, the heard voice is represented as having many properties distinct from the subject's own voice, and this suggests a clear difference in phenomenology between inner speech and AVH. Thus, if AVH is misattributed inner speech, self-monitoring accounts must explain this transformation, specifically the acquisition of properties normally associated with auditory experience of another voice. Failure of self-monitoring does not explain that.

Seal et al. (2004) propose that once inner speech output is generated, top-down cognitive processes including expectations, response biases and delusions mislead the patient to take the output as another's voice (p.65). AVH results from

---

<sup>13</sup> An n-value is not reported in the paper for this, but given the other values reported, it is likely in the mid-40s.

the patient's misinterpretation of an ambiguous inner speech output. Yet patients consistently attest to AVH as involving properties found in *auditory* experience of someone else's voice. While top-down processes might explain misattributing identity, I am doubtful that they can satisfactorily explain the auditory phenomenology of AVH such as heard differences in location, pitch, timbre, accent and intensity. These doubts based on phenomenology do not refute the mechanism, but they weigh against it and suggest an alternative answer to the identification question as a serious competitor: AVH is (aberrant) auditory experience that represents the properties in question.<sup>14</sup>

Along these lines, self-monitoring accounts can posit a more complicated mechanism that renders inner speech into something more clearly auditory in format. Let us assume there is such a mechanism. Its output would be the generation of an auditory experience. We then return to the two 'crude' processes associated with the strong and moderate theses noted above, and the following question arises: why opt for the more complicated process where a mechanism first generates inner speech experiences that are then transformed to AVH when the postulated auditory output could be produced directly? A moderate mechanism avoids an inner speech intermediary as unnecessary and thus, postulates no inner speech episode to lose track of. It is compatible with this moderate picture that auditory mechanisms involve some of the mechanisms generating normal inner speech. Indeed, imaging work comparing inner speech against auditory imagery of voices show activation of many of the same brain areas (Shergill et al., 2000).

---

<sup>14</sup> Another challenge to inner speech self-monitoring models is that some patients *answer their voices by inner speech* (12/29 patients in Langdon et al., 2009)

Thus, a simpler answer to the identification question circumvents the strong thesis: AVH is just auditory experience of another voice, one whose production might call upon some inner speech associated mechanisms. Self-monitoring accounts can endorse this answer: AVH arises when patients lose track of active auditory experience (e.g. auditory imagery), not inner speech. Such auditory episodes, being experienced as spontaneous, would then be experienced as actual auditory experience of a voice (something like this view is suggested by Hoffman et al., 2007, p. 1172). Notice, however, that an alternative presents itself: rather than being active, the auditory states might be automatic or involuntary. After all, consider when a tune pops into your head. Such episodes are felt as involuntary happenings. There is some feature of the normal brain that yields such automatic auditory states, and this suggests a different hypothesis, one that cannot be ruled out at this juncture: since in mundane experience, there are such automatic experiences, AVH arises when the basis of normal automaticity goes awry. That is, AVH amounts to the persistent, automatic generation of a specific form of auditory experience.

In speaking of auditory experience, I am intentionally vague about whether the state is like auditory perception, auditory imagery (imagination), or auditory memory (the latter two need not be mutually exclusive, though imagery may involve generation of new representations). In the end, the account I favor can encompass all or a subset of these; what it emphasizes is the *automaticity* of such auditory states as a basis of AVH. There are distinctions to be drawn between these types of auditory states that may correlate with differences in phenomenology and content, but for present purposes, it is enough to note that they may have different neural underpinnings. For example, auditory

recollection (Wheeler et al., (2000)) or auditory imagery of another voice (see Shergill et al. 2001) might activate secondary rather than primary auditory cortex while auditory perceptual states will involve the latter as well. The auditory underpinning of AVH is supported by imaging data that frequently shows activation of primary and/or secondary auditory cortex during AVH (see reviews in Allen et al. (2008) and Jadri et al. (2010); recall the diversity of imaging results noted in the previous section). I shall explain the underlying notion of automaticity in section 5, but the current point is that the phenomenology and some imaging data provide evidential support for identification of AVH with auditory experience of a voice.

What of empirical evidence that suggests a role for inner speech in AVH? I cannot review this evidence here, but the distinction between the strong and moderate theses matters. Specifically, imaging data showing similar brain activity in inner speech and in AVH supports only the moderate thesis, namely that the phenomena share some mechanisms or have co-localized mechanisms (see Allen et al. and Jadri et al., *op. cit.*). Accordingly, such correlations are compatible with the claim that AVH is aberrant auditory experience that can arise, in the cases at issue, in part from inner speech mechanisms but without inner speech episodes as intermediates. I conclude that there is a strong *prima facie* case that self-monitoring accounts, by appealing to the strong thesis, give the wrong answer to the identification question. Let us now turn to the phenomenological question and alienness.

#### **4. Alienness and AVH**

Talk of 'alienness' while compelling as a description does not provide a clear explanatory target. The required clarity is achieved by linking alienness to representational content, and this can explain why alienness arises without invoking self-monitoring. In addition, I argue that self-monitoring makes the wrong predictions in respect of alienness.

What is alienness? The term likely covers a wide range of features, but in the absence of a clear account, I propose a characterization in terms of 'otherness' and 'alienation', two features that are tied to AVH content as follows: (A) in AVH, the patient experiences *another's* voice (thus, in that sense an alien/other), a voice located in the head or in the external environment where there is no such voice; (B) often, the voice says things contrary to what the speaker thinks, and in this way, the subject is *alienated* from the voice.

Let's begin with (A), which is the crucial feature found in nearly all cases of AVH.<sup>15</sup> Consider some typical properties of AVH. First, the onset of AVH seems not to be under the agent's control. AVH captures the subject's attention. Second, AVHs have properties that are characteristic of auditory perception, namely they involve auditory representations of a voice that registers its timbre, pitch, intensity, distinctive semantic content and identity. These features are characteristic of normal auditory experiences of another's voice, for in typical cases, a voice intrudes on our experience and captures our attention where this sound is auditorily represented to us as having distinctive properties of pitch, timbre, intensity, and semantic content. Where we have such experiences of

---

<sup>15</sup> One caveat: there are a few cases of patients who report hearing their own voices (see Hoffman et al., 2007). We need to know more about the total content of these episodes, but it would be odd to hear one's voice when one knows one is not otherwise speaking. In that way, one's own voice can have an 'other' quality to it.

someone else speaking that recur, then we would have something odd. The presence of this 'other' voice in AVH experience is a central feature in the experience of alienness (otherness, externalization).

A contributing feature, in many cases, is that what the voice expresses is incongruent with the patient's views. Moritz and Larøi (2008) report that among normal subjects who hear voices, 67% of them report that the voice's utterances reflect their own thoughts whereas this drops to 16% in subjects with schizophrenia; in contrast, 41% of the latter report that the thoughts do not at all reflect their thoughts while only 17% of normal subjects note this. Hoffman et al. (2007) also report that 47% of their patient population (n=44) report the voices are incongruent with their thought most to all of the time (nearly 70% of the subjects report incongruency as being often; around 95% mention that it is incongruent at least sometimes). This incongruency, when it occurs, plausibly contributes to a sense of alienation and in that way may contribute to a sense of alienness.

Once we have provided this initial characterization of what alienness is, we have a simple explanation of why it occurs in AVH: alienness occurs in AVH because of the typical *content* of AVH. AVH is a type of auditory experience (auditory hallucination) that represents a distinct voice which does not exist at that time and whose utterances are often incongruent with the patient's own views. Certainly, where these episodes persist, we have something that is worthy of the description *alien*. We can thus answer the phenomenological question without appeal to self-monitoring. The explanation also dispels some of the obscurity behind talk of alienness, but I do not claim to have captured everything



that patients and clinicians may refer to by it. I urge theorists who invoke the notion, however, to be clearer as to what they mean by it.

What of self-monitoring explanations of alienness? The basic problem is that self-monitoring accounts get the structure of the phenomenology wrong. Recall the distinction between episode and content. In AVH we have an experiential event that also represents a voice (sound) as present when in fact there is no such voice. On the self-monitoring account, what goes wrong is that the agent loses track of the *event* and attributes it to an external, alien source. This common description implies that the patient is confused about the source of an internal episode, attributing *its* source to someone else. Thus, Stephens and Graham (2003), enunciating a version of the strong thesis, characterize alienness as follows: the agent ‘experiences *the episode* as alien—that is, as somehow attributable to another person rather than to the subject’ (4, my emphasis)). Yet this is precisely not what patients seem to experience. Their attention is not directed at an internal episode but rather at a putative voice (sound). All that is required for the attributed externalization is that patients have an auditory experience that represents the presence of another’s voice. There is no need for the subject to take the episode into account; the *content* of the episode does the phenomenological work.

We bring this out by showing that self-monitoring accounts make an incorrect prediction. Recall self-monitoring explanations of delusions of control and thought insertion. In delusion of control, the patient loses track of the action event: when a bodily movement is made, self-monitoring fails to tag the movement as self-generated. The patient then attributes ‘other’ agency to the episode (e.g., *this action* is brought about by someone else). In the case of thought

insertion, the subject loses track of a thought and attributes the episode as belonging to another (e.g., *this thought* belongs to someone else). Since the form of explanation is *episode* misattribution, the analog in AVH should be auditory *insertion*: *this auditory experience* is another's experience not the patient's. But as far as I know, this has never been reported: patients do not claim to have someone else's auditory experience.<sup>16</sup> It is not the episode that is externalized but the voice as represented by the episode (first sense of 'voice'). I conclude that self-monitoring gives the wrong answer to the phenomenological question because it generates the wrong prediction.

What precisely do self-monitoring accounts explain in respect of AVH? They either provide incorrect answers to or are not needed to answer the identification and phenomenological questions. The last of the core features noted above is the involuntariness of AVH. Since this is a feature of the *episode*, self-monitoring does have a plausible explanation: AVH seems involuntary precisely because the episode is not tagged as self-generated. I noted, however, that there is an alternative namely the automatic generation of auditory states that would accordingly also feel involuntary. I want to clarify what this alternative comes to by explaining the notion of automaticity. While I think the incorrect phenomenological prediction weighs against self-monitoring accounts of AVH, deciding which account is correct requires empirical work elucidating mechanisms. It will be progress, however, to define clearly an alternative.

## 5. Automaticity and AVH

---

<sup>16</sup> There does seem to be a case of 'perception broadcast' (Schmid-Siegel et al. 2004). Thanks to Ian Gold for pointing this out.

The notions of *automaticity* and the correlated notion of *control* are widely deployed in psychology, yet they are not well defined. In this section, I am going to define one notion of automaticity that I will use to characterize AVH. I begin with an intuitive notion of agentic control and define automaticity in contrast to it. I shall then define the notion of passivity in terms of automaticity, and suggest that in respect of episodes of AVH, the subject is passive.

Agentic control is exemplified in a subject's intentional behavior. Automaticity, in contrast, is exemplified in those features that are not intentionally brought about but simply happen. The account of automaticity and control will involve two ideas. The first is the notion of top-down modulation, a notion that is available once we represent a cognitive system as having a hierarchical structure and exhibiting causal feedback.<sup>17</sup> This suggests an intuitive conception of control in terms of a certain type of causal feedback from higher-level systems. The relevant influence will be top-down control by the subject's intentions. The second is that whatever automaticity and control come to, they are *relative* to specific features of the process in question. That is to say, a given cognitive process may have certain properties that are subject to control and certain properties that are not and hence automatic. To give an intuitive case: that I move my arm rather than my leg or that I think about my mother rather than my father identify features of the process (moving an arm, thinking about my mother) that are exactly what I intend to bring about and hence control. But it

---

<sup>17</sup> A major question is how to construct such hierarchies in a principled way. Still, hierarchical characterizations are common in cognitive science and for the specific form of top-down control, namely intentional control, the top-down description is likely accurate across different accounts of hierarchy.

may also be the case that my thinking of this particular image or moving my arm with this particular speed or defined trajectory is not controlled but merely automatic. What this shows is that a cognitive or motor process can exemplify properties that are controlled and properties that are not. This *relativization* is crucial to an adequate conception of both notions.

Let us begin with the following account of agentive control in respect of some cognitive process of type *P* undertaken by a subject *S* with respect to a specific property *F* of that process:

(AC) A token process, *X*, of type *P* is *agentively controlled* by *S* with respect to its having property *F* if and only if *X*'s having *F* is the result of top-down modulation of a specific sort due to *S*, namely from *S*'s intention that there be a *P*-type process having *F*.<sup>18</sup>

So, a specific movement of the subject's arm has the property of being a controlled reaching for a mug iff *S* reaches for the mug as a result of intending to do so. Automaticity is the absence of control in this sense. In this case, we can speak of agentive automaticity as follows:

(AA) A token process, *X*, of type *P* is *agentively automatic* relative to subject *S* with respect to its having *F* if and only if *X*'s having *F* is not the result of

---

<sup>18</sup> We need a clause to rule out causal deviance. I shall sidestep that issue though I assume such a clause can be given, a perennial problem in philosophy of action. We can also broaden the focus to a broader category of motivational states including desires and emotions.

top-down modulation of a specific sort due to *S*, namely from *S*'s intention that there be a *P*-type process having *F*.

Thus, the observed kinematics of reaching movements that are consistent across subjects are not features of the movement that the subject intends to bring about (she doesn't intend to reach with such kinematic features, namely one characterized by an earlier acceleration and then deceleration phase, see Jeannerod, (1988)). Hence, the presence of these features is automatic. In this way, the agent's reaching for the mug has both automatic and controlled features. I do not claim that these notions capture everything that psychologist's mean in thinking of control and automaticity although I do think it defines one form, what we might call intentional control. This is, of course, the control we associate with agency.

Finally, in the case where none of the features of a process are subject to agentic control, we have passivity:

(P) *S* is *passive* in respect of token process, *X*, of type *P* if and only if for every property *F* of *X*, *X*'s having *F* is not the result of top-down modulation of a specific sort due to *S*, namely from intentions that there be a *P*-type process having *F*.

That is, *S* is passive in respect of a process *X* just in case every property of that process is automatic. Thus, *S* does not exert any control over *X*.

What is distinctive about many of the features of schizophrenia, including AVH, is that the agent is seemingly passive with respect to them in the sense just

defined. Notice that self-monitoring accounts assume that AVH is the result of intentional control where self-monitoring goes awry. I propose an alternative: AVH is genuine passivity as defined. The properties of AVH are not subject to the agent's control and thus not something she intentionally brings about. If my alternative is correct, an explanation of AVH will require that we understand its automaticity. The contrast with self-monitoring accounts then is stark: is AVH a result of an action control system that fails or is it genuinely automatic?

If AVH episodes are automatic auditory states, they will be generally experienced as involuntary and not under direct control. Of course, automaticity on its own is not sufficient for psychopathology as the normal cases of tunes popping into the head demonstrate. Another critical feature may be the persistence of AVH. While it might not be too disturbing to have a single involuntary image of one's dead friend's voice berating one for some embarrassing episode in one's past, it would be disturbing should such automatic states persistently arise. Given persistent automaticity, the level of disruption and disturbance would increase. The issue then is not just automaticity but *persistent* automaticity. Once we have a voice that has the representational properties noted above and are persistently automatic, we have a symptom more recognizable as AVH. Still, none of this need deny the contribution of the sort of cognitive factors and propensities noted by e.g. Seal et al. (2004). The full story will likely involve many factors even if involuntariness in AVH has its roots in automaticity and not failure of self-monitoring.

If my proposal is correct, we should expect to find evidence of over-activation in auditory sensory areas in patients with AVH. Stephane et al. (2001) review changes in the molecular properties of cells in the superior temporal

gyrus (STG, this includes primary auditory cortex) that ‘could reflect a functional overactivation of the speech perception area’, (p. 63); they note studies that show correlation between changes in STG volume and propensity to have auditory hallucinations (on this see also Allen et al., 2008, p. 178-9); and they discuss cases of auditory hallucinations that arise from brain damage where lesions are almost always found in sensory areas. Several authors in reviews of recent empirical work emphasize overactivity or failure of inhibition in sensory areas (Allen et al, 2008; Jardi et al., 2010). Accordingly, the automaticity of auditory experience account proposed here deserves further scrutiny.

What of empirical evidence supporting a role for self-monitoring, specifically cases that show decrements or absence of activity in regions implicated in self-monitoring (Shergill et al., 2000 and 2001)?<sup>19</sup> A systematic discussion of the empirical literature cannot be done here, but the observed result is compatible with my account. The central point is this: *both automaticity and self-monitoring models predict the failure (absence) of self-monitoring in AVH with the corresponding decrement or absence of activity in brains regions involved in self-monitoring processing.* This feature is built into the self-monitoring account of AVH as part of a posited defect in the action-control system. However, on my account, we also expect a decrease or absence of activity in regions implicated in self-monitoring since the episodes are generated automatically and this means

---

<sup>19</sup> An early PET imaging study suggested that patients with schizophrenia show lower levels of self-monitoring than control groups and patients who don’t suffer AVH (McGuire et al., 1996). Subjects were asked to imagine another voice speaking sentences where these sentences were earlier generated by the subject via their own inner speech. Only those patients with ongoing AVH showed a decrease in activation relative to controls in a region of the supplementary motor area associated with self-monitoring. This suggested a decrement in self-monitoring of inner speech. Still, the patient’s intentional generation of auditory imagery *was not experienced as an episode of AVH*, something one expects given a defect in self-monitoring.

that they are outside the action-monitoring system (the system is not built to monitor events that just happen). Clearly, what is needed is work that goes beyond correlation and more directly addresses causation and mechanisms.<sup>20</sup>

I conclude that self-monitoring accounts fail to answer or incorrectly answer certain core questions regarding AVH and in that respect, fail to explain schizophrenia. The proper locus of explanation for self-monitoring accounts vis-à-vis AVH is the sense of involuntariness, but we have an alternative explanation where defects in self-monitoring play no explanatory role. The automaticity account I have proposed then answers our opening questions as follows: AVH is an automatic auditory experience of another's voice represented as having specific audible properties where ascribed alienness is explained by the content of that experience and where involuntariness is explained by automaticity. The crucial question, and in many ways the central question, remains: why does AVH have that specific range of reported content? We have yet, I think, a fully satisfying answer on that score. Nevertheless, the automaticity account provides a promising beginning of an explanation of AVH and hence of schizophrenia.

## **6. Positive Symptoms: A Unified Causal Basis?**

One lesson that I hope emerges is that we need to pay more careful attention to AVH phenomenology. Indeed, it may be the case that there will not be a single

---

<sup>20</sup> Asai et al. (2008) report a strong correlation between self-monitoring deficits (as measured in a motor task) and propensity to auditory hallucination in their pool of subjects with schizotypal features, a propensity to certain symptoms of schizophrenia. This is a suggestive report in a non-clinical population although the authors caution that the results are preliminary. Resolution, again, requires that we probe the causal mechanisms.



mechanism that explains all instances of AVH in schizophrenia (as Larøi (2006) and Jones (2010) argue). I want to close with a similar possibility that has been obscured by the dominance of self-monitoring accounts in explaining positive symptoms. *If* it were the case that self-monitoring accounts do adequately explain delusions of control and that AVH is driven in many cases by automaticity, this would mean that there are two mechanisms leading to a sense of involuntariness found in the positive symptoms. Specifically, the boundary between these will depend on whether the symptoms are ‘agentive’ or not. First, in delusions of control, we would have a failure of action monitoring. Some feature of normal action-control, self-monitoring, is defective, and here, schizophrenia involves a defect in agency. Let us categorize such positive symptoms as *agentive symptoms*. Second, in AVH, I have argued in favor of an automaticity account of involuntariness where involuntariness is defined in terms of automaticity, i.e. the absence of agentive control. In this case we have what we might call a *non-agentive symptom*. Positive symptoms may thus divide along an agency/non-agency (control/automaticity) divide.

Whether such a divide is exemplified in the positive symptoms is, of course, an empirical question. To the extent that understanding the causal mechanisms facilitates identifying new treatments, we must investigate the possibility that two common positive symptoms have different causal bases. This means that we will need to ask questions that reflect this distinction and perhaps undertake different experimental approaches. Only then will we have a chance in fully explaining schizophrenia.

*Center for the Neural Basis of Cognition  
Carnegie Mellon University*

## References

- Aggernaes, A. 1972: The difference between the experienced reality of hallucinations in young drug abusers and schizophrenic patients. *Acta Psychiatrica Scandinavica*, 48, 287-299.
- Aleman, A. et al. 2003: Cognitive basis of hallucinations in schizophrenia: role of top-down information processing. *Schizophrenia Research*, 64, 175-185.
- Allen, P., Aleman, A. & McGuire, P.K. 2007: Inner speech models of auditory verbal hallucinations: Evidence from behavioural and neuroimaging studies. *International Review of Psychiatry*, 19, 409-17.
- Allen, P. et al. 2008: The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neuroscience and Biobehavioral Reviews*, 32, 175-191.
- Alpert, M. & Silvers, K.N. 1970: Perceptual characteristics distinguishing auditory hallucinations in schizophrenia and acute alcoholic psychoses. *American Journal of Psychiatry*, 127, 298-302.
- Asai, T. & Sugimori, E. & Tanno, Y. 2008: Schizotypal personality traits and prediction of one's own movements in motor control: What causes an abnormal sense of agency? *Consciousness and Cognition*, 17, 1131-42.
- Baker, C.A. & Morrison, A.P. 1998: Cognitive processes in auditory hallucinations: attributional biases and metacognition. *Psychological Medicine*, 28, 1199-1208.
- Bentall, R., Haddock, G. & Slade, P. 1994: Cognitive behavior therapy for persistent auditory hallucinations: from theory to therapy. *Behavior Therapy*, 25, 51-66.
- Blakemore, S.J., Wolpert, D. & Frith, C. 2000: Why can't you tickle yourself? *Neuroreport*, 11, R11-16.
- Brébion, G. et al. 2000: Positive symptomatology and source-monitoring failure in schizophrenia: an analysis of symptom-specific effects. *Psychiatry Research*, 95, 119-31.
- Cho, R. & Carter, C. In press: Monitoring Deficits in Schizophrenia. In D. Barch, ed. *Handbook of Cognitive and Affective Neuroscience of Psychopathology*. New York: Oxford University Press.
- Copolov, D.L. et al. 2003: Cortical activation associated with the experience of auditory hallucinations and perception of human speech in schizophrenia: a PET correlation study. *Psychiatry Research: Neuroimaging*, 122, 139-152.

- David, A.S. 2004: The cognitive neuropsychiatry of auditory verbal hallucinations: an overview. *Cognitive Neuropsychiatry*, 9, 107-123.
- Ditman, T. & Kuperberg, G.R. 2005: A source-monitoring account of auditory verbal hallucinations in patients with schizophrenia. *Harvard Review of Psychiatry*, 13, 280-299.
- Evans, C.L., McGuire, P.K. & David, A.S. 2000: Is auditory imagery defective in patients with auditory hallucinations? *Psychological Medicine*, 30, 137-148.
- Feinberg, I. 1978: Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin*, 4, 636-40.
- Franck, N. et al. 2000: Confusion between silent and overt reading in schizophrenia. *Schizophrenia research*, 41, 357-64.
- Frith, C. & Done, D. 1988: Towards a neuropsychology of schizophrenia. *The British Journal of Psychiatry*, 153, 437-43.
- Frith, C., Blakemore, S. & Wolpert, D. 2000: Explaining the symptoms of schizophrenia: Abnormalities in the awareness of action. *Brain Research Reviews*, 31, 357-63.
- Frith, C. 1995: *The Cognitive Neuropsychology Of Schizophrenia* New edition., Psychology Press.
- Garrett, M. & Silva, R. 2003: Auditory hallucinations, source monitoring, and the belief that "voices" are real. *Schizophrenia Bulletin*, 29, 445-457.
- Hoffman, R. E. 2009: Verbal Hallucinations and Language Production Processes in Schizophrenia. *The Behavioral and Brain Sciences*, 9, 503-48.
- Hoffman, R. E. et al. 2007: Experiential features used by patients with schizophrenia to differentiate 'voices' from ordinary verbal thought. *Psychological Medicine*, 38, 1167-1176.
- Hurlburt, R.T. 1990: *Sampling Normal and Schizophrenic Inner Experience*, Plenum Press.
- Hurlburt, R.T. & Schwitzgebel, E. 2007: *Describing Inner Experience?: Proponent Meets Skeptic*, MIT Press.
- Jardri, R. et al. 2010: Cortical Activations During Auditory Verbal Hallucinations in Schizophrenia: A Coordinate-Based Meta-Analysis. *The American Journal of Psychiatry*.
- Jeannerod, M. 1988: *The Neural and Behavioural Organization of Goal-Directed Movements*, Oxford: Oxford University Press.

- Johns, L.C. et al. 2001: Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine*, 31, 705-715.
- Jones, S. 2010: Do we need multiple models of auditory verbal hallucinations? Examining the phenomenological fit of cognitive and neurological models. *Schizophrenia Bulletin*, 36, 566-75.
- Jones, S.R. & Fernyhough, C. 2007: Thought as action: Inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition*, 16, 391-399.
- Keefe, R.S. et al. 2002: Source-monitoring deficits for self-generated stimuli in schizophrenia: multinomial modeling of data from three sources. *Schizophrenia Research*, 57, 51-67.
- Langdon, R. et al. 2009: The phenomenology of inner speech: comparison of schizophrenia patients with auditory verbal hallucinations and healthy controls. *Psychological Medicine*, 39, 655-663.
- Larøi, F. 2006: The phenomenological diversity of hallucinations: Some theoretical and clinical implications. *Psychologica Belgica*, 46, 163-83.
- Leudar, I. et al. 1997: What voices can do with words: pragmatics of verbal hallucinations. *Psychological Medicine*, 27, 885-898.
- MacKay, D.G. 1992: Constraints on theories of inner speech. *Auditory imagery*, 121-149.
- McGuire, P.K. et al. 1996: The neural correlates of inner speech and auditory verbal imagery in schizophrenia: relationship to auditory verbal hallucinations. *The British Journal of Psychiatry: The Journal of Mental Science*, 169, 148-159.
- McGuire, P.K., Murray, R.M. & Shah, G.M.S. 1993: Increased blood flow in Broca's area during auditory hallucinations in schizophrenia. *The Lancet*, 342, 703-706.
- Moritz, S. & Larøi, F. 2008: Differences and similarities in the sensory and cognitive signatures of voice-hearing, intrusions and thoughts. *Schizophrenia research*, 102, 96-107.
- Nayani, T.H. & David, A.S. 1996: The auditory hallucination: a phenomenological survey. *Psychological Medicine*, 26, 177-189.
- Schmid-Siegel, B., Stompe, T. & Ortwein-Swoboda, G. 2004: Being a webcam. *Psychopathology*, 37, 84-5
- Seal, M.L., Aleman, A. & McGuire, P.K. 2004: Compelling imagery, unanticipated

- speech and deceptive memory: neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry*, 9, 43-72.
- Shergill, S. et al. 2000: Functional anatomy of auditory verbal imagery in schizophrenic patients with auditory hallucinations. *The American Journal of Psychiatry*, 157, 1691-1693.
- Shergill, S., Bullmore, E. & Brammer, M. 2001: A functional study of auditory verbal imagery. *Psychological medicine*, 31, 241-53.
- Stephane, M., Barton, S. & Boutros, N.N. 2001: Auditory verbal hallucinations and dysfunction of the neural substrates of speech. *Schizophrenia Research*, 50, 61-78.
- Stephane, M. et al. 2003: The internal structure of the phenomenology of auditory verbal hallucinations. *Schizophrenia Research*, 61(2-3), 185-193.
- Stephens, G.L. & Graham, G. 2003: *When Self-Consciousness Breaks*, MIT Press.
- van de Ven, V.G. et al. 2005: The spatiotemporal pattern of auditory cortical responses during verbal hallucinations. *Neuroimage*, 27, 644-655.
- Wheeler, M.E., Petersen, S.E. & Buckner, R.L. 2000: Memory's echo: vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11125.
- Wolpert, D. & Ghahramani, Z. 2000: Computational principles of movement neuroscience. *Nat Neurosci*, 3 Suppl, 1212-7.