Between Scientific and Empathetic Understanding: The Case of Auditory Verbal Hallucination

Forthcoming in British Journal for the Philosophy of Science

Please cite published version

Shivam Patel

Abstract

A common but overlooked form of explanation in psychiatry is what I label 'empathetic explanation'. Empathetic explanations invoke empathetic variables, which, in addition to providing an explanation of the target phenomenon, also afford an empathetic understanding of it. Focusing on the case of auditory verbal hallucination (AVH), I argue that empathetic explanation fails to provide an adequate account of the phenomenon, perniciously shapes empirical research, and confuses empathetic understanding with scientific understanding. I close by providing a general condition on the legitimate application of empathetic explanation.

1. Introduction

Explanations in psychiatry attempt to provide a broad and comprehensive understanding of a target system by drawing on a number of different explanatory levels (Kendler, 2012). Consider visual hallucination. Suppose we know that visual hallucination is caused by excess dopamine absorption in the striatum; that the right anterior insula is strongly implicated; and that hallucination is tightly correlated with childhood abuse and trauma. That is, imagine we have a complete account of visual hallucination in terms of molecular, neural, and social factors. Do we thereby have the sort of broad and comprehensive understanding sought by a science of mental illness? Are molecular, neural, and social explanations sufficient for a comprehensive scientific grasp of visual hallucination? Many would say not. At least one crucial level of explanation seems to be missing – cognitive explanation (Fodor [1968]; Fletcher and Frith [2009]).

The target of this paper is a subclass of cognitive explanation often used to explain mental illness, which I call 'empathetic explanation'. Empathetic explanations explain an experience or behaviour in terms that also provide one with an empathetic grasp of that experience or behaviour. Not all cognitive explanations are empathetic explanations. The cognitive explanation of speech production appeals to top-down connections from semantic to motor levels (Levelt [1993]), while face perception is explained in terms of a representation of the difference between a given face and an average face (Tsao and Livingstone [2008]). These are cognitive explanations, but not empathetic ones, since the explanations do not include variables that provide us with empathetic understanding of the targets. For example, knowing that speech production implicates top-down connections from semantic to motor levels gives us no sense of what it is like to produce speech. Conversely, the phenomenology of speech production would not render its cognitive explanation at all intuitive. In contrast, empathetic explanations are cognitive explanations that involve

empathetic variables, variables that provide empathetic understanding. Consider the following cognitive explanation of Mary's purchase of ice cream: Mary bought ice cream because she had a craving for ice cream. In addition to being a cognitive explanation of Mary's ice cream purchase, this explanation also puts us 'in Mary's shoes': we know what it is like to buy something in the grip of a craving. Empathetic explanation is often implicit in intentional explanation of everyday behaviour (Goldman [1989]).

Such explanations have also creeped into scientific domains. My focus will be on empathetic explanations of the first-rank symptoms of people diagnosed with schizophrenia, including delusion and hallucination. Explanations of thought insertion appeal to the actual possession of thoughts (Ratcliffe and Wilkinson [2015]), explanations of delusion often appeal to states of imagination or belief (Currie and Ravenscroft [2002]; Bortolotti [2010]), and explanations of auditory verbal hallucinations routinely cite aberrations of inner speech or auditory imagery (Waters et al. [2006]; Moseley and Wilkinson [2014]). In each of these cases, the relevant symptom is explained in terms of mental states found in the population not suffering from the symptom. This appeal to 'normal' mental state kinds puts the first-rank symptoms within a certain kind of empathetic reach: thought insertion can be empathetically understood in terms of what it would be like to have one's own thoughts inserted into one's head; delusion in terms of what it would be like to experience the recalcitrance of one's own irrational beliefs or imaginings; and auditory verbal hallucination in terms of what it would be like to experience certain aberrations of one's own inner speech or auditory imagery. Unlike with the aforementioned cognitive explanations of speech production and face perception, we gain an empathetic understanding of thought insertion, delusion, and hallucination by appeal to the relevant empathetic variables.

Kendler and Campbell ([2014]) are recent proponents of such explanations, claiming that '[a] functional level of analysis…provide[s] the rubric for developing an understanding of the experiences of illness' (p. 6). The authors illustrate their claim by appeal to delusions of reference. Delusions of reference involve mistakenly taking neutral events to be especially significant and to carry important messages concerning oneself (Startup et al. [2009]). For example, watching a news anchor on TV, someone may sense that the anchor is directing her report at them and imbue it with hidden meaning. Kendler and Campbell offer a neuropsychological explanation of delusions of reference in terms of aberrations in salience processing. According to this explanation, in delusions of reference the salience system misfires and mistakenly tags as salient otherwise mundane items in the environment (see (Kapur [2003])). This functional explanation of delusions of reference is supposed to provide empathetic understanding of the target phenomenon. Consider Kendler and Campbell:

> Assume we were at some time in the future when it has been well established that the DA salience system is dysfunctional in schizophrenia. You are interviewing Mr. X as he describes this experience [delusions of reference] to you. Will you…find this entirely un-understandable, as a 'quite alien mode of experience'? Or, in the context of your knowledge of the pathophysiology of schizophrenia, will this be understandable?...We suggest that the latter will be more likely. Given our expanded knowledge of neuroscience, we can convert these experiences of the intrusion of meaning into consciousness from the un-understandable to the understandable. (p. 3)

According to Kendler and Campbell, the functional explanation contains an element – the misattribution of salience – that makes delusions of reference understandable. The idea is that since we have 'all experienced a misattribution of meaning' (for example, mistaking a stranger for a loved one) we can put this experience of misattribution together with incoming stimuli (for example, hearing a reporter) and thereby empathetically understand delusions of reference. Kendler and Campbell are optimistic about empathetic explanations, claiming that they 'now permit us to expand the boundaries of the understandable in psychiatric illness'. They compare

this expansion to the way in which 'the microscope and the telescope have allowed us to peer into previously unknown worlds' (p. 2).

Against Kendler and Campbell, I contend that empathetic explanations of the first-rank symptoms of schizophrenia are illegitimate. The idea that empathetic explanations are potentially problematic is not new. One lesson from Nagel's discussion of what it is like to be a bat is that it is a mistake to assume that there must be an empathetic variable that explains the phenomenal goings-on of the bat (Nagel [1974]). Jaspers, in his classic work *General Psychopathology*, emphasizes the importance of distinguishing between 'explanation' (*Erklären*) and 'understanding' (*Verstehen*) in the domain of mental illness. According to Jaspers, our causal explanations of mental illness should not appeal to empathetic variables. Jaspers ([2007], p. 176) goes so far as to say that causal explanation and empathetic understanding 'do not impinge on one another and neither has the right to criticize the other since both pursue entirely different aims'. Finally, Hempel – as different in temperament from Jaspers and Nagel as one can be – also holds that '[e]mpathic insight and subjective understanding provide no warrant or objective validity, no basis for the systematic prediction or explanation of specific phenomena…' ([1970], p. 163).

My particular focus in this paper will be on auditory verbal hallucination (AVH). AVH is the experience of perceiving a voice in a situation in which there is no actual voice to be heard (DSM-5 [2013], p. 87).[1] I think research into AVH has failed to appreciate the potential pitfalls of empathetic explanation. In Section 2, I describe empathetic explanations of AVH. Section 3 then elucidates underlying requirements on such explanations: the matching condition, the preservation condition, and the abnormal phenomenology condition. Section 4 shows that

[1] The discussion to follow will cover two kinds of AVH: 'outer space' and 'inner space' varieties (Ratcliffe [2017]). Whereas 'outer space' AVH involve hearing an auditory voice as present in external space, 'inner space' AVH involve 'hearing' a non-auditory voice or registering a pure meaning in one's own head.

empathetic explanation of AVH fails because none of these conditions are satisfied, while Section 5 responds to an attempt to rescue such explanation on the basis of empirical evidence. In Section 6, I diagnose the attraction and the error of empathetic explanation. And finally, in Section 7, I close by tempering Kendler and Campbell's optimism that neurofunctional explanations can provide for empathetic understanding of symptoms of psychosis.

2. Characterizing Empathetic Explanation of AVH

There are currently several competing cognitive explanations of AVH. Most of these explanations are empathetic insofar as the explanations appeal to variables that afford an empathetic understanding of the symptom. The most popular posits inner speech as the basis of AVH (see, for example, Frith [1992]; Seal [2004]; Jones and Fernyhough [2007]; Swiney and Sousa [2014]). In order to understand this explanation, it is important to understand the mechanisms underlying online speech production. Online speech production is believed to involve both feedforward and lateral processes. The feedforward process starts with an intention to speak that generates speech motor commands, which are then executed by the vocal tract. Along with this feedforward process, there is also a lateral process whereby a copy of the speech motor commands is sent to a 'forward model', which generates a prediction of the auditory consequences of executing those commands. This prediction is supposed to serve as a check on the vocal tract movement: if there is a mismatch between the predicted auditory consequence and the actual auditory signal, the system computes the difference as error, and makes corrections (see, for example, Pickering and Garrod [2013]).

Parts of this process are thought to be re-used in the production of inner speech (see, for example, Swiney and Sousa [2014]; Carruthers [2018]). On this view, one produces an intention to speak, which generates speech motor commands, but, in the case of inner speech, one suppresses those commands at the vocal tract. Despite this suppression, the lateral process is still engaged and a prediction of auditory consequence is generated. Inner speech, on this view, is identified with the prediction of auditory consequence: hearing one's voice in one's head.

Building on this model of inner speech, many authors argue that in auditory verbal hallucination, there is a mismatch between the original speech intention and the prediction of auditory consequence (Figure 1) (see, for example, Swiney and Sousa [2014]).

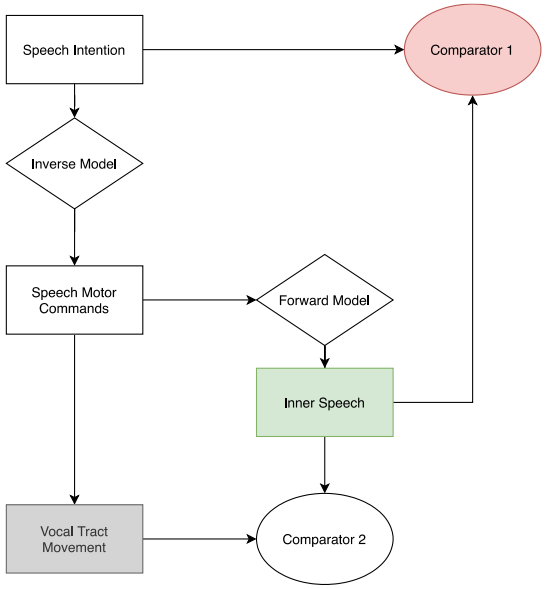

Figure 1: A partial model of auditory verbal hallucination in terms of simplified speech control mechanisms

To use a toy example, one may generate the intention to say 'The book!', but the prediction may turn out to be 'Send backup!'. This mismatch is supposed to give rise to 'an unusual feeling of agency…[that] inner speech is outside of intentional control' (Swiney and Sousa [2014], p. 11). Because people with the diagnosis of schizophrenia often misattribute self-generated information to others, it is hypothesized that they misattribute inner speech to an external agent. As a result,

they experience an external agent saying, 'Send backup!'. On this view, AVH amounts to the experience of inner speech as 'alien', or otherwise as produced by some external agent.[2]

Despite their popularity, inner speech models of AVH have come under attack on phenomenological grounds (Wu [2012]; Cho and Wu [2013]). Wu ([2012], p. 96) cites phenomenological reports suggesting that while AVH is often auditory in character, inner speech is 'more often abstracted from an auditory format'. In addition, according to Cho and Wu ([2013]), inner speech is most often in one's own voice, while AVH is in the voice of another. In light of these differences in phenomenology, Wu ([2012], p. 94) claims that in order for the predictive mechanism proposed by inner speech theorists to be viable it must be equipped with 'an additional mechanism to effect the transformation' from a non-auditory state (inner speech) to a state with auditory phenomenology (AVH). In effect, according to Cho and Wu, since inner speech does not have the phenomenology of AVH, inner speech theorists must posit a process that transforms the phenomenology of inner speech into that of AVH. However, Wu ([2012], p. 97) asks, '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?'. In short, according to Wu, inner speech does not explain AVH because inner speech does not possess a phenomenology that matches that of AVH.

As an alternative, Cho and Wu suggest that a more plausible view is that the basis of AVH is auditory imagery. Unlike inner speech, auditory imagery is often in the voice of another person

[2] This is not the only model of AVH that implicates inner speech. According to an earlier model, AVH is generated when inner speech fails to be predicted (Seal et al. [2004]). On a more recent model, inspired by predictive processing, AVH is caused when the prediction error generated by inner speech fails to be suppressed (Wilkinson and Fernyhough [2017]). These models differ in their details, but are united in representing inner speech as a distal cause of AVH and prediction error as a more proximate cause. Moreover, according to all three models, AVH consists in the misattribution of inner speech. Thus, in addition to being a cause of AVH, according to these models, inner speech also partially constitutes the phenomenon. Though I target a particular inner speech model here, my criticisms of it focus on assumptions that unite all three (see Section 3).

and is, by definition, auditory. On this alternative, AVH amounts to auditory imagery that is both involuntary and persistent. According to Wu, the automatic and persistent activation of auditory imagery would cause the sort of 'disruption and disturbance' indicative of AVH (Wu [2012]).

Inner speech theorists have responded to Cho and Wu by arguing that inner speech does match the phenomenology of AVH. Consider Moseley and Wilkinson ([2013]):

> Cho and Wu simply assume that inner speech is always experienced in one's own voice, and are not aware of research suggesting that the presence of other people's voices is exactly the kind of quality reported in typical inner speech. (p. 1)

According to inner speech theorists, since inner speech is auditory in character and involves the voice of other people, there is no need for 'an additional mechanism' to transform the inner speech into a state whose phenomenological properties match those of AVH. Subsequent debate between the two camps has centered on whether inner speech or auditory imagery better matches the phenomenological properties of AVH (see, for example, Cho and Wu [2014]).

A central assumption guiding the debate between inner speech and auditory imagery theorists is that explanations of AVH must cite a normal mental state kind that, in some sense, matches the phenomenology of an AVH. On the one hand, Cho and Wu claim that inner speech cannot explain AVH because inner speech fails to match the phenomenological properties of AVH. Cho and Wu then offer auditory imagery as a more plausible substrate for explaining AVH. Though neither Cho nor Wu end up accepting the auditory imagery model, they claim that its closeness in phenomenology to AVH makes auditory imagery a better explanatory variable than inner speech.[3] On the other hand, inner speech theorists respond to Cho and Wu by affirming that inner speech is a close match to the phenomenology of AVH. We can make sense of this response

[3] Cho and Wu ([2013]) offer an account of AVH in terms of the spontaneous activation of brain areas involved in representing the voice of another person. However, it is not clear how this is an informative, cognitive-level explanation. In the first instance, AVH just is the representation of a voice when there is no actual voice. To then say, as Cho and Wu do, that the representation of a voice is caused by brain areas involved in representing voices is no longer to provide a distinctively cognitive-level explanation of AVH.

only if we assume that inner speech theorists are also presupposing that explanations of AVH must cite a mental state variable that is, in some sense, a phenomenological match with AVH. This assumption is thus serving as common ground and guiding debate between the two parties.

3. Conditions on Empathetic Explanation of AVH

The assumption of a phenomenological match suggests that inner speech and auditory imagery theorists both offer empathetic explanations of AVH. In both cases, one can draw on one's own inner speech or auditory imagery to grasp the phenomenological characteristics of AVH: AVH is what it would be like for me to feel that my inner speech was not produced by me or to have involuntary auditory imagery. Echoing Kendler and Campbell, since we have 'all experienced' various coerced/involuntary actions (for example, a doctor unexpectedly moving our arm) we can put these experiences together with our own inner speech or auditory imagery to empathetically understand AVH. In order to get more precise on the specific kind of empathetic explanation on offer, I will now describe three implicit, interlocking conditions of such explanation: the matching condition, the preservation condition, and the abnormal phenomenology condition.

First the matching condition. Inner speech and auditory imagery theorists both assume that AVH is to be explained in terms of a mental state that matches the phenomenology of AVH. But in what sense does the phenomenology of the mental state variable need to match that of AVH? In order to understand this we need to distinguish two parts of the phenomenology of an AVH episode: one part that is shared between the mental state kind and the hallucination – call this the 'standard phenomenology' – and another part that is unique to AVH and distinguishes it from the

mental state kind – call this the 'abnormal phenomenology'. This brings us to the matching condition: for any explanation of some AVH A, where a normal mental state kind N is part of the explanans, N possesses all and only the standard phenomenological properties possessed by A, but not its abnormal phenomenology.

The second condition follows from the first. The aberrant process that operates over the mental state variable to generate an AVH must preserve the standard phenomenology of the mental state. After all, if the standard phenomenology were not preserved, then the mental state and the AVH would not match in their standard phenomenology. This condition is supposed to be satisfied by the mechanisms offered by both inner speech and auditory imagery theorists. For inner speech theorists, the mismatch at the comparator brings about the feeling of other-authorship of inner speech but does not change its standard phenomenology. For auditory imagery theorists, too, the aberrant process – persistent automaticity – does not change the standard phenomenology of the auditory imagery. The preservation condition states that for any explanation of some AVH A in terms of some normal mental state kind N, there is an abnormal process over N that preserves all and only the standard phenomenology of N.

The third condition follows from the first two: that the AVH possesses an abnormal phenomenology that distinguishes it from the mental state variable. According to Cho and Wu, the abnormal phenomenology is a feeling of automaticity or involuntariness, while for inner speech theorists, it is a feeling of other-authorship or alienness. Thus, while preserving standard phenomenology, the aberrant process also bestows abnormal phenomenology onto the AVH. The abnormal phenomenology condition states that for any explanation of some AVH A, the explanans must explain the presence of an abnormal phenomenology in A.

The form of empathetic explanation offered by inner speech and auditory imagery theorists is characterized in terms of the conjunction of the matching, preservation, and abnormal phenomenology conditions. More generally, an explanation of AVH puts it in empathetic reach insofar as it appeals to a normal mental state kind that matches the standard phenomenology of AVH (matching condition) as well as to a process whereby it comes to possess an abnormal phenomenology (preservation and abnormal phenomenology conditions). Of course, the precise shape of an empathetic explanation will differ for different target phenomena, but these three conditions capture empathetic explanation of AVH.

4. Against Empathetic Explanation of AVH

I will now argue that explanations that conform to these conditions are problematic.

4.1 Against the Matching Condition

Even the most promising candidate mental state kinds fail to match the standard phenomenology of AVH. My argument for this will proceed in two steps, addressing two types of AVH: auditory and non-auditory (Ratcliffe [2017]).

The paradigmatic type of AVH is auditory. Consider the following first-person report: 'Most of the time I can hear it like it was just someone standing next to me. It's a different feeling than when you think words inside of your head…' (Woods et al. [2015], p. 326). I will assume (with imagery theorists) that auditory imagery is the best candidate to match the auditory type of

AVH. However, there is converging evidence that auditory imagery fails to match the life-like auditory features present in AVH.

First, whereas auditory AVH is often reported to possess a particular level of loudness (Vercammen, et al. [2010]), auditory imagery in general does not represent loudness (Pitt and Crowder [1992]; Bishop et al. [2013]). Pitt and Crowder ([1992]) show that perceived tones matching in loudness generate priming effects, whereas imagined tones do not, suggesting that unlike auditory perception, auditory imagery does not represent loudness. Second, whereas AVH possesses a life-like tempo, the tempo of auditory imagery is degraded (Janata and Paroo [2006]). Halpern ([1988]) found that subjects often 'level out' differences in tempo in imagery: subjects slowed the tempo of a fast song and speeded the tempo of a slow song. Third, AVH is often reported as representing the precise vocal timbre of some particular agent (Junginger and Frame [1985]). However, using the same paradigm as their loudness experiment, Pitt and Crowder ([1992]) concluded that auditory imagery fails to represent dynamic timbre (see also Bailes [2007]). Taken together, the matching condition seems to fail for auditory imagery: auditory imagery often does not possess standard phenomenological features present in auditory AVH, and when it does, it is not life-like.[4]

The second important subtype of verbal hallucination is non-auditory. Consider the following report: 'I did not hear the voices aurally. They were much more intimate than that, and inescapable. It's hard to describe how I could 'hear' a voice that wasn't auditory…' (Woods et al.

[4] An imagery theorist may object that I have only cited studies that employ healthy subjects. According to the objection, imagery in those suffering AVH might be more life-like than in healthy subjects. Although subjective reports have suggested that those suffering AVH have more acute and intense imagery (Mintz and Alpert [1972]), objective measures show that subjects in fact have the same imagery abilities (Aleman and Haan [1998]). These objective measures include paradigms in which subjects judge the oddball from imagined sounds or judge stress on the basis of imagined lyrics, among other behavioral tasks (Aleman, et al. [2000]). The fact that subjects with a diagnosis of schizophrenia do not differ from controls on such measures suggests that the above findings are also applicable to subjects with a diagnosis.

[2015], p. 326). The best candidate for explaining non-auditory AVH would presumably be some form of abstract, non-auditory inner speech. In order to argue that such inner speech fails to fulfil the matching condition, I will target a difference in the linguistic contents of inner speech and AVH.

Tovar et al. ([2019]) transcribed the verbal content of AVH as subjects experienced it and found that first-person pronouns were present in only twelve percent of reports, while second- and third-person pronouns were present in forty-eight and forty-four percent of reports, respectively – a pattern that is not present in the inner speech of people with AVH (Langdon et al. [2009]). In addition, Hoffman et al. ([2008], p. 1170) found that '46 percent of respondents reported that verbal content of voices [in AVH] was distinct from verbal thought [inner speech] either most of the time or all the time'. Overall, then, the phenomenological literature on inner speech suggests that the matching condition fails: although inner speech is often experienced as in the first-person, AVH is not so experienced. (See Wu [2012] for a similar line of reasoning against inner speech theorists.) The upshot is that neither auditory imagery nor inner speech matches the standard phenomenology of AVH.

One might object that the psychological variable in an empathetic explanation must match AVH only with regard to determinable standard phenomenological properties. On this view, a given substrate needs to match only with respect to such determinable properties as being auditory, but not specific determinate properties, for example, having such and such timbre. This would allow auditory imagery, which matches the determinable properties of AVH, to serve as a psychological variable in the explanation of AVH.

However, this response is not available to either inner speech or auditory imagery theorists. This is because the aberrant processes that they posit do not involve any transformation of their

respective psychological variables. On the one hand, a mismatch between intention and prediction does not change the linguistic content of the prediction, and on the other, automatic and persistent auditory imagery does not transform the phenomenology of the auditory imagery itself. Thus, both models require that the mental state variable and AVH match in their maximally determinate standard phenomenology. Of course, theorists may posit a process that transforms a state with determinable properties into an AVH with certain determinate properties. However, the process would be hopelessly *ad hoc*: it would be a process that converts determinable properties into determinate properties possessed by AVH, leaves unchanged any determinate properties shared with AVH, and cancels any determinate properties not shared with AVH. Better to give up the matching condition than engage in such backflips.

Another objection is that I have shown only that inner speech and auditory imagery fail to match the standard phenomenology of AVH. But, according to the objection, this does not entail that the matching condition itself is mistaken. The objector grants that inner speech and auditory imagery fail to satisfy the matching condition, but maintains that cognitive explanations of AVH must be given in terms of a normal mental state kind that matches the standard phenomenology of AVH. However, if inner speech and auditory imagery are the best candidates to match the standard phenomenology of AVH, then, if these states fail, there is plausibly no normal mental state that satisfies the matching condition. But this suggests that we should give up the matching condition – it would be strange to hold on to the condition if it is known not to be satisfied by even the best candidates. Therefore, even though the argument does not directly target the matching condition, it does so indirectly by targeting the only plausible states that might satisfy it.

4.2 Against the Preservation Condition

According to the preservation condition, the transformation over some mental state variable must preserve its standard phenomenology. Both inner speech and auditory imagery theorists attempt to satisfy the preservation condition by appealing to the unintended activation of either inner speech or auditory imagery. I will argue that neither the inner speech nor the auditory imagery theorist's way of explaining the notion of unintended activation provides us with a tenable account of AVH. This failure stems from an allegiance to the preservation condition, and so I will conclude that the preservation condition should be rejected.

For the auditory imagery theorist, unintended activation occurs when an intention fails to control or modulate a mental state. The activation of auditory imagery is automatic, according to Wu, in the sense that there is no feature of the auditory imagery that is the result of top-down modulation. However, according to Wu, automaticity of auditory states is itself insufficient to generate AVH, since cases of having a tune pop into your head are also automatic. Wu adds that the auditory state must also be persistent in the sense that it is activated repeatedly. The problem for this view is that obsessive-compulsive disorder (OCD) also often involves the automatic and repetitive activation of visual and auditory imagery (Lipton et al. [2010]). For example, Brown ([2006]) describes a person (himself) with a perpetual music track running in their head throughout the day. Though the perpetual track is disruptive and disturbing, it is not reported to be an AVH. Thus, the process posited by Wu to account for AVH is insufficient to generate it. Rather, the process at best accounts for the repetitive and uncontrollable imagery we observe in OCD.

According to inner speech theorists, unintended activation amounts to a mismatch between an intention and prediction at a comparator. On this model, a match at the relevant comparator initiates a feeling of agency, whereas a mismatch does not (Swiney and Sousa [2014]) (see Figure

1). The problem for this view is that the same mismatch is also implicated in a far more mundane phenomenon, covert repair, whereby subjects catch themselves before making a speech error. On standard speech control models, covert repair is thought to be subserved by a mismatch between intended speech and predicted speech (Postma and Kolk [1993]). This mismatch between intention and prediction does not seem to generate a feeling of loss of agency of the robust sort posited in explanations of AVH. In fact, it is quite the opposite: a condition on engaging in covert repair is that the speech that is predicted but not intended is recognized as a mistake made by oneself. This makes it implausible that people with a diagnosis of schizophrenia will misattribute the prediction despite the fact that they tend to misattribute self-produced information. Thus, the mismatch posited by inner speech theorists is insufficient to generate AVH. Rather, the process at best accounts for covert repair.

The processes proposed by inner speech and auditory imagery theorists are insufficient to generate AVH. Instead, the processes explain a mental state that is relatively more mundane: OCD, in the one case, and covert repair, in the other. As a result, if inner speech and auditory imagery are the best candidate mental state kinds, then it seems wrong to think that the process that generates AVH must be a process that preserves the standard phenomenology of inner speech or auditory imagery.

One might object that I have shown only that the processes of mismatch and spontaneous activation fail to generate AVH. This does not entail that the preservation condition itself is mistaken. The objector might grant that the two processes fail to generate AVH, while still maintaining that cognitive explanations of AVH must be given in terms of a process that preserves the standard phenomenology of a normal mental state kind. However, that it is unclear which other processes could be responsible for AVH once we select inner speech or auditory imagery as

the normal mental state kind. Inner speech is characterized in terms of its functional role in the comparator architecture, and so it is not clear that there is some other comparator-specific breakdown that could be a candidate to generate AVH. And though auditory imagery is less tied down to some specific architecture, it is hard to see how the imagery theorist can do any better than appeal to spontaneous activation to account for AVH. Now, if inner speech and auditory imagery are the best candidate normal mental state kinds, and the processes in question are the only plausible candidates, then we have an argument that the preservation condition cannot be satisfied without falling short of an account of AVH. The preservation condition should be rejected because there are no plausible AVH-generating processes that satisfy it.

4.3 Against the Abnormal Phenomenology Condition

The third and final condition is that empathetic explanations of AVH explain how AVH comes to possess an abnormal phenomenology that distinguishes it from the normal mental state kind. Most theorists believe that the abnormal phenomenology of AVH at least involves the feeling of a lack of control. This putative abnormal phenomenology is implied by a number of different descriptions, including 'alienness', 'extraneity', 'other-authorship', 'automaticity', 'spontaneity', and the like. These quotes express this dominant view:

> [the] proposal is then made that top-down factors…lead to unintended inner speech being experienced as *other-generated* (italics added) (McCarthy-Jones [2012], p. 251)

> …all accounts of AVH must explain (A) the spontaneity of AVH episodes (they often just happen). Both accounts can deal with the spontaneity: in self-monitoring, it is explained by spontaneous failure of self-monitoring so that AVH feels spontaneous; in the spontaneous activity account, it is the actual spontaneous activity of auditory areas (Cho and Wu [2013], p. 4)

Auditory imagery theorists capture the feeling of a lack of control in terms of the feeling of passivity and automaticity, while inner speech theorists capture it in terms of a feeling of alienness or other-authorship.

We can show what is wrong with the abnormal phenomenology condition by first considering the following reports of OCD and anarchic hand syndrome, respectively:

> I began to feel I was being taken over by some force in my brain that I could not control, until eventually my days and nights became ruled by its orders: 'If you don't wash your hands twenty times, you will have to kill yourself . . . Oh no, I don't think you got the soap underneath every fingernail on that nineteenth wash, so it's all negated, you must start the washes again . . .['] (National Collaborating Centre for Mental Health [2006], p. 52)

> while watching television, [a woman] noted her left hand flinging across her visual field. Her left hand stroked her face and hair without her will. She got terrified. Her attempts to control the left hand with the right hand were unsuccessful. (Panikkath et al. [2014], p. 219)

The sufferer of OCD does not want to think thoughts about hand washing and suicide but has those thoughts anyway. Similarly, the patient with anarchic hand syndrome reports her inability to control the movements of her left hand. A feeling of a lack of control is thus cited as part of the abnormal phenomenology of these conditions. What is felt to go wrong has to do with the inability to control thoughts or movements, respectively.

Contrast these reports with those of AVH, where we do not find reports of a lack of control. The following are samples taken from an online forum for people diagnosed with schizophrenia:

> To me the voices always sound like someone is saying something to me just outside the room i'm [*sic*] in or having a conversation just outside. Usually it is 3 voices. One constantly says he wants to kill me or hurt me in some way [*sic*]. The other two can go from trying to talk him down or to saying they want to kill me too. (from SP2342, *Schizophrenia.com*[5])

> I only had auditory hallucinations twice…The other time I was on the street and it was a guy calling my name, I thought it were some random dudes on [*sic*] a coffee shop and ran away from them scared shitless. (from anon72351231, *Schizophrenia.com*[6])

Similar reports are found in scientific literature investigating the phenomenology of AVH:

---

[5] Retrieved on August 20, 2022 from https://forum.schizophrenia.com/t/people-with-auditory-hallucinations-question/53126/19

[6] Retrieved on August 20, 2022 from https://forum.schizophrenia.com/t/people-with-auditory-hallucinations-question/53126/19

> Starting when I was about 20 years old, I heard the voices of demons screaming at me, telling me that I was damned, that God hated me, and that I was going to hell... (Woods et al. [2015], p. 326)

> I hear a mixture of men and women, but no children. They usually tell me to do things, but not dangerous things. Like they'll tell me to take out the garbage or check the lock on the window or call someone. (Woods et al. [2015], p. 325)

Unlike the reports of OCD and anarchic hand syndrome, these reports of AVH do not cite a feeling of a lack of control. Instead, the reports involve hearing people talking outside the room, hearing someone calling one's name, hearing screaming, and hearing a voice telling one to take out the garbage.

What determines whether or not the feeling of a lack of control is cited as part of the abnormal phenomenology of these conditions? Take first the case of OCD and anarchic hand syndrome. A feeling of a lack of control shows up as abnormal in these conditions precisely because thoughts, on the one hand, and limb movements, on the other, are normally the sorts of things that are under some kind of control. My hand movements are normally under my control, and so when my hand moves in the absence of that control, the lack of control is experienced as abnormal. Although the way in which we are in control of our thoughts is perhaps more difficult to pinpoint, when we experience thoughts outside of our control, we experience them as abnormal because thoughts are (in some sense) normally under our control.

Why, then, do reports of AVH not cite a feeling of a lack of control? The above reports of AVH are reports of the experience of speech perception – hearing people talking outside the room, hearing screaming, and so on. Unlike thought and movement, speech experience *normally* involves either a feeling of a lack of control or is neutral with respect to feeling control or feeling its lack. After all, in the normal case I cannot control what I hear you say nor can I control my hearing what you say. At most I might manipulate your responses via conversational scheming (in the first instance) or cup my ears (in the second instance), but these manipulations do not

provide the sense of control at issue here. Thus, if AVH is a species of speech experience, then lack of control enters into AVH as a normal phenomenological factor, not as an extra, abnormal phenomenological factor. I suggest that the abnormal phenomenology condition should be rejected.[7]

A number of studies show that those who can control the frequency and/or onset of AVH episodes have more overall well-being than those with no such control (for example, Baumeister [2017]; Swyer and Powers [2020]). An objector might argue that, since lack of control reduces well-being, lack of control is what is experienced as abnormal in AVH. However, the finding does not suggest this conclusion. Compare the case of panic attacks. Gaining control of panic attacks may lead to an increase in overall well-being, but this would tell us nothing about the phenomenology of panic attacks themselves (for example, the sense of impending doom, racing heartbeat, and so on). A similar point seems to hold for AVH. That AVH episodes come and go outside the scope of one's control does not tell us about what is experienced as abnormal about hearing a voice where there is no voice to be heard.

Of course, there is something experienced as abnormal about AVH. Inner speech and auditory imagery theorists assume that the abnormal phenomenology of AVH is an intrinsic feature of the experience – a feeling of other-authorship or automaticity. Assuming that the abnormal phenomenology is intrinsic enables empathetic explanation by allowing the theorist to mark the difference between the normal mental state kind and AVH. But if we reject the framework of empathetic explanation, and the need to appeal to a normal mental state kind in the first place, we put ourselves in a position to provide a much simpler account of the abnormal phenomenology. I

[7] To be clear, I am not denying that the feeling of a lack of control is part of the phenomenology of AVH, but just that it is part of its abnormal phenomenology. Reports that do mention the feeling of a lack of control – for example, 'when you hear…shouting, it's like you can't intercept, so my thoughts can't override it…' (Foxwell [2022], p. 75) – should accordingly be interpreted as reporting on the normal phenomenology of speech perception.

suggest that what is centrally experienced as abnormal about AVH is just that it involves hearing a voice where there is no voice to be heard. On this alternative, the abnormal phenomenology of AVH is not intrinsic, but relational: one experiences a voice in a context in which there is no voice to be heard.

5. Rescuing Empathetic Explanation of AVH?

We have considered two attempts to provide empathetic explanations of AVH, one in terms of inner speech, the other in terms of auditory imagery. Inner speech and auditory imagery theories allow one to gain an empathetic understanding of AVH (Section 3), but fail to deliver a plausible explanation of the phenomenon (Section 4). Appeal to either inner speech or auditory imagery gives rise to a form of empathetic explanation whose conditions – the matching, preservation, and abnormal phenomenology conditions – fail to be satisfied. The central lesson is that there is little positive reason to think that explanations featuring empathetic variables succeed in explaining AVH.

However, one might attempt to rescue empathetic explanation of AVH by appeal to neuroimaging. Authors have inferred that inner speech is causally implicated in AVH based in part on similarities in neural activation between inner speech and AVH, for example, in Broca's area. A clear instance of this sort of inference comes from McGuire, et al. ([1996], p. 36):

> The finding that both inner speech and auditory verbal imagery were associated with increased activity in Broca's area is of particular interest, as increases in this region have been identified in schizophrenics experiencing auditory verbal hallucinations (McGuire et al. 1993). This suggests that the silent articulation of verbal thoughts, which is common to both processes, may be involved in the generation of verbal hallucinations.

This same inference is present in a review of the AVH literature by Allen, Aleman, and McGuire ([2007], pp. 412-13), who state that 'the observed activity in Broca's area during AVH implicates the involvement of inner speech and/or auditory verbal imagery'. These authors seem to assume that observing mental states with overlapping neural activations licenses the inference that one of the mental states causes or underlies the other mental state.

However, we can see what is wrong with this inference once we note that there is overlap in the brain processes implicated in X-ing (for example, walking) and imagining X-ing (for example, imagining walking) (Jeannerod [2006]). But this alone should not lead us to infer that imagining X-ing is part of an explanation of X-ing (or vice versa). For example, we should not assume that imagining walking is part of an explanation of walking just on the basis of the overlap in processing. Likewise for the relationship between inner speech and auditory verbal imagery, on the one hand, and AVH: the mere fact that there is an overlap in activation does not justify the inference that one is explanatory of the other. The inference would appear to be a good one only if it is already being assumed that there must be a normal mental state kind that underlies AVH. Only then would matching activation constitute good evidence that inner speech and/or auditory imagery explains AVH. But this would involve assuming, in advance of any empirical inquiry, a framework of explanation that this paper has argued is problematic.

6. Empathetic Explanation: Attractions and Potential Pitfalls

Empathetic explanation fails to provide an adequate account of AVH. Not only are there compelling arguments against such explanation (Section 4), but empirical discussion in its favour mistakenly presuppose empathetic variables (Section 5). But if empathetic explanations of AVH

are misguided, we face a puzzle: why have theorists in both philosophy and psychology been attracted to them?

6.1. The Attraction

Kendler and Campbell explain the appeal through a narrative. They have us imagine that researchers have discovered a causal relationship between certain DNA variants and a particular psychiatric symptom, say, hallucination. Though this finding provides some understanding of how hallucination arises, Kendler and Campbell emphasize that stopping the explanation here is 'clearly unsatisfactory', since all that has been found is 'the start of a causal chain'. Kendler and Campbell then have us imagine that researchers fill in the causal pathway in terms of both microanatomical dysfunctions involving neurotransmitters and more general dysfunctions in network dynamics. For Kendler and Campbell this is 'still not adequate', since there remains the leftover question: how do we get from the 'brain-based set of findings' to the 'symptoms' of the disorder? The same sentiment is also expressed in Fletcher and Frith ([2009], p. 49):

> although it is relatively easy to understand how a brain disorder might cause a loss of sensation or an inability to think, it is more difficult to understand how a brain disorder can create new and compelling experiences. Such explanations require one to forge a link between brain activity and the subjective experience of a mind. Explanations such as 'hallucinations are caused by overactive dopamine receptors' are unsatisfactory because they leave an explanatory gap between the mental and the physical. How can dopamine cause a voice or a belief?

These authors all pose the same question. In Fletcher and Frith's terms, how can a researcher close the 'gap' between 'brain activity' and 'subjective experience'? In Kendler and Campbell's terms, 'how do we get from…brain-based set of findings…to the symptoms of a disorder?'[8] Both pairs

[8] We should not interpret these authors as invoking the hard problem of consciousness. Given that the authors believe that an appeal to cognitive/functional explanation is an answer to their questions, charity demands a more modest interpretation.

agree that the gap is to be closed by appeal to cognitive, functional, or neuropsychological explanation (see Section 1).

But not just any cognitive, functional, or neuropsychological explanation will work. For example, if we explain speech production in terms of top-down connections from semantic to motor levels, we do not gain any insight into the subjective experience of speech production. Or if we explain face perception in terms of a representation of the difference between a given face and an average face, we do not understand the subjective experience of face perception. Where these cognitive explanations fail is exactly where empathetic explanation is supposed to do its work. According to Kendler and Campbell, such explanation serves as a bridge between 'brain-based findings' and the 'subjective experience' of psychiatric illness, by 'translating' the former into the register of the latter:

> The neuropsychological expansion of understanding relies on more than the functional analysis of brain functioning…. It relies on the fact that neuropsychology provides hypotheses about the functional and physiological correlates of familiar subjective experiences, and thus puts us in a position to understand the significance of disturbances in those underlying structures. (Kendler and Campbell [2009], p. 5)

The idea is that by correlating our own subjective experience with parts of a functional explanation, we are able to gain an understanding of otherwise bizarre experiences. The inner speech and auditory imagery accounts of AVH work in just this way: AVH is presented as being just like what would happen to me if my auditory imagery were automatic and persistent (imagery theorists) or as being just like what would happen to *me* if *my* inner speech were to not feel self-produced (inner speech theorists). In both cases, we use our own experiences of inner speech or auditory imagery, lack of control, and so on, to come to an understanding of AVH via the functional explanation. Thus, the central attraction of empathetic explanation – cognitive explanation with empathetic variables – is that it allows one to understand the subjective experience of psychiatric illness, and

thereby comprehend how brain language ‘translates’ into mind language (see also (Gough [forthcoming]) for discussion).

6.2. The Potential Pitfall

However, empathetic explanations are problematic, at least when it comes to AVH. This is in contrast to the case of Mary’s ice cream purchase. This gives rise to a more general question: what marks the distinction between the legitimate use of empathetic explanation, observed in the explanation of Mary’s ice cream purchase, and its illegitimate use, observed in the case of AVH? In explaining why Mary bought ice cream, we appeal to craving. The evidence we have for the explanation includes, for example, the fact that craving causes people to purchase ice cream, and that Mary’s behaviour is indicative of craving, and so on. What makes the explanation of Mary’s ice cream purchase a legitimate application of empathetic explanation is that the plausibility of the explanation does not rest on the fact that it provides empathetic understanding of Mary’s behaviour. Instead, the explanation is supported by empathy-independent evidence, with empathetic understanding being at most incidental with respect to that evidence.

In contrast, in the illegitimate use of empathetic explanation observed in the case of AVH, empathetic understanding is not incidental, but itself constitutes part of the support for offering the empathetic explanation in the first place. The imagery theorist objects to the inner speech model on the ground that inner speech does not match the standard phenomenology of AVH, while the inner speech theorist responds by affirming that they do match. Whether a given variable counts as explanatory of AVH is supposed to turn on whether it is a phenomenological match with AVH (see Section 2). The error of empathetic explanations of AVH is that there are no grounds for

positing inner speech or auditory imagery as causal variables other than the fact that these states provide us with empathetic understanding.

7. Conclusion: Tempering Optimism about Empathetic Explanation

The paper offers a cautionary tale: beware of cognitive explanations that include empathetic variables, since there is a risk that the variables provide empathetic understanding without genuine explanation. This cautious attitude is a counterweight to Kendler and Campbell's optimism about empathetic explanation. I agree with Kendler and Campbell that there may well be explanations of the neurofunctional sort that could afford one an empathetic understanding. What this paper points out, however, is that there is a risk, borne out in the case of AVH, of going in for an empathetic explanation not because there is evidence that a given empathetic variable explains a set of symptoms, but simply because it gives us an empathetic grasp of the target phenomenon.

Acknowledgements

Many thanks to Edouard Machery, Wayne Wu, James Shaw, Peter Langland-Hassan, Zina Ward, Zachary Gabor, Savannah Leon, and audiences at Florida State University, the 2021 Pacific Division meeting of the American Philosophical Association, the 2021 meeting of the Southern Society for Philosophy and Psychology, and the 2021 meeting of the Association for the Advancement of Philosophy and Psychiatry.

Department of Philosophy
Florida State University
Tallahassee, United States of America
smpatel@fsu.edu

References

Aleman, A. and E. H. de Haan. [1998]: 'On Redefining Hallucination', *The American Journal of Orthopsychiatry* **68**, pp. 656–59.

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. [2000]: 'Mental Imagery and Perception in Hallucination-prone Individuals', *The Journal of Nervous and Mental Disease*, **188**, pp. 830–36.

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**, pp. 407–15.

American Psychiatric Association. [2013]: *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5*, Washington, D.C: American Psychiatric Publishing.

Bailes, F. [2007]: 'The Prevalence and Nature of Imagined Music in the Everyday Lives of Music Students', *Psychology of Music*, **35**, pp. 555–70.

Baumeister, D., Sedgwick, O., Howes, O., & Peters, E. [2017]: 'Auditory verbal hallucinations and continuum models of psychosis: A systematic review of the healthy voice-hearer literature', *Clinical Psychology Review*, **51**, pp. 125–41.

Bishop, L., Freya B., and Roger T. [2013]: 'Musical Imagery and the Planning of Dynamics and Articulation During Performance', *Music Perception: An Interdisciplinary Journal*, **31**, pp. 97–117.

Bortolotti, L. [2010]: *Delusions and Other Irrational Beliefs*. New York: Oxford University Press.

Brown, S. [2006]: 'The Perpetual Music Track', *Journal of Consciousness Studies*, **13**, pp. 25–44.

Campbell, J. [1999]: 'Schizophrenia, the Space of Reasons, and Thinking as a Motor Process', *The Monist*, **82**, pp. 609–25.

Carruthers, P. [2018]: 'The Causes and Contents of Inner Speech', in P. Langland-Hassan and A. Vicente (*eds*.), *Inner Speech: New Voices*. New York: Oxford University Press, pp 31-52.

Cho, R. and Wu, W. [2013]: 'Mechanisms of Auditory Verbal Hallucination in Schizophrenia', *Frontiers in Psychiatry*, **4**, pp. 1-8.

———. [2014]: 'Is Inner Speech the Basis of Auditory Verbal Hallucination in Schizophrenia?', *Frontiers in Psychiatry*, **5**, pp. 1-3.

Currie, G. and Ravenscroft, I. [2002]: *Recreative Minds: Imagination in Philosophy and Psychology*. New York: Oxford University Press.

Fletcher, P. and Frith, C. [2009]: 'Perceiving Is Believing: A Bayesian Approach to Explaining the Positive Symptoms of Schizophrenia', *Nature Reviews. Neuroscience*, **10**, 48–58.

Fodor, J. [1968]: *Psychological Explanation: An Introduction to the Philosophy of Psychology*. New York: Random House.

Foxwell, J. [2022]: 'Lost Agency and the Sense of Control', in A. Woods, B. Alderson-Day, & C. Fernyhough (*eds*.), *Voices in Psychosis: Interdisciplinary Perspectives*. New York: Oxford University Press, pp. 74-81.

Frith, C. [1992]: *The Cognitive Neuropsychology of Schizophrenia*. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Goldman, A. I. [1989]: 'Interpretation Psychologized', *Mind and Language*, **4**, pp. 161–85.

Gough, J. [forthcoming]: 'On the Proper Epistemology of the Mental in Psychiatry: What's the Point of Understanding and Explaining?', *The British Journal for the Philosophy of Science*.

Halpern, A. [1988]: 'Perceived and Imagined Tempos of Familiar Songs', *Music Perception: An Interdisciplinary Journal*, **6**, pp. 193–202.

Hempel, C. G. [1970]: *Aspects of Scientific Explanation: And Other Essays in the Philosophy of Science*. New York: Free Press.

Hoffman, R. E., 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**, pp. 1167–76.

Janata, P. and Paroo, K. [2006]: 'Acuity of Auditory Images in Pitch and Time', *Perception & Psychophysics*, **68**, pp. 829–44.

Jaspers, K. [1963]: *Karl Jaspers' General Psychopathology*. Chicago: The University of Chicago Press.

———. [2007]: 'Causal and Understandable Relationships between Events and Psychosis in Dementia Praecox (Schizophrenia)', in H. Sass (*ed*.), *Anthology of German Psychiatric Texts*, New Jersey: John Wiley & Sons. pp. 174-293

Jeannerod, M. [2006]: *Motor Cognition: What Actions Tell the Self*. New York: Oxford University Press.

Jones, S. and Fernyhough, C. [2007]: 'Thought as Action: Inner Speech, Self-Monitoring, and Auditory Verbal Hallucinations', *Consciousness and Cognition*, **16**, pp. 391–99.

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**, pp. 149–55.

Kapur, S. [2003]: 'Psychosis as a State of Aberrant Salience: A Framework Linking Biology, Phenomenology, and Pharmacology in Schizophrenia', *The American Journal of Psychiatry*, **160**, pp. 13–23.

Kendler, K. [2012]: 'The Dappled Nature of Causes of Psychiatric Illness: Replacing the Organic-Functional/Hardware-Software Dichotomy with Empirically Based Pluralism', *Molecular Psychiatry*, **17**, pp. 377–88.

Kendler, K. and J. Campbell. [2014]: 'Expanding the Domain of the Understandable in Psychiatric Illness: An Updating of the Jasperian Framework of Explanation and Understanding', *Psychological Medicine*, **44**, pp. 1–7.

Langdon, R., Jones, S., Connaughton, E., & Fernyhough, C. [2009]: 'The phenomenology of inner speech: Comparison of schizophrenia patients with auditory verbal hallucinations and healthy controls', *Psychological Medicine*, **39**, pp. 655–63.

Levelt, W. J. M. [1993]: *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.

Lipton, M. G., Brewin, C. R., Linke, S., & Halperin, J. [2010]: 'Distinguishing features of intrusive images in obsessive-compulsive disorder', *Journal of Anxiety Disorders*, **24**, pp. 816–22.

McCarthy-Jones, S. [2012]: *Hearing Voices: The Histories, Causes and Meanings of Auditory Verbal Hallucinations*. Cambridge: Cambridge University Press.

McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. [1996]: 'Functional anatomy of inner speech and auditory verbal imagery', *Psychological Medicine*, **26**, pp. 29–38.

Mintz, S. and M. Alpert. [1972]: 'Imagery Vividness, Reality Testing, and Schizophrenic Hallucinations', *Journal of Abnormal Psychology*, **79**, pp. 310–16.

Moseley, P. and Sam W. [2014]: 'Inner Speech Is Not so Simple: A Commentary on Cho and Wu (2013)', *Frontiers in Psychiatry*, **5**, pp. 1-2.

Nagel, T. [1974]: 'What Is It Like to Be a Bat?', *The Philosophical Review*, **83**, pp. 435–50.

National Collaborating Centre for Mental Health (UK). [2006]: *Obsessive-Compulsive Disorder: Core Interventions in the Treatment of Obsessive-Compulsive Disorder and Body Dysmorphic Disorder*. United Kingdom: British Psychological Society.

Panikkath, R., Panikkath D., Mojumder D., and Nugent K. [2014]: 'The Alien Hand Syndrome', *Baylor University Medical Center Proceedings*, **27**, pp. 219–20.

Pickering, J. and Garrod, S. [2013]: 'An Integrated Theory of Language Production and Comprehension', *The Behavioral and Brain Sciences*, **36**, pp. 329–47.

Pitt, M. A., and R. G. Crowder. [1992]: 'The Role of Spectral and Dynamic Cues in Imagery for Musical Timbre', *Journal of Experimental Psychology: Human Perception and Performance*, **18**, pp. 728–38.

Postma A. and Kolk H. [1993]: 'The Covert Repair Hypothesis', *Journal of Speech, Language, and Hearing Research*, **36**, pp. 472–87.

Ratcliffe, M. and Wilkinson, S. [2015]: 'Thought Insertion Clarified', *Journal of Consciousness Studies: Controversies in Science & the Humanities*, **22**, pp. 246–69.

Ratcliffe, M. [2017]: *Real Hallucinations: Psychiatric Illness, Intentionality, and the Interpersonal World*. Cambridge, MA: MIT Press.

Seal, M., Aleman, A., & McGuire, P. [2004]: 'Compelling Imagery, Unanticipated Speech and Deceptive Memory: Neurocognitive Models of Auditory Verbal Hallucinations in Schizophrenia', *Cognitive Neuropsychiatry*, **9**, pp. 43–72.

Startup, M., Bucci, S., & Langdon, R. [2009]: 'Delusions of Reference: A New Theoretical Model', *Cognitive Neuropsychiatry*, **14**, pp. 110–26.

Swiney, L. and Sousa P. [2014]: 'A New Comparator Account of Auditory Verbal Hallucinations: How Motor Prediction Can Plausibly Contribute to the Sense of Agency for Inner Speech', *Frontiers in Human Neuroscience*, **8**, pp. 1-15.

Swyer, A., and Powers, A. R. [2020]: 'Voluntary Control of Auditory Hallucinations: Phenomenology to Therapeutic Implications', *Schizophrenia*, **6**, pp. 1-9.

Tovar, A., Fuentes-Claramonte P., Soler-Vidal J., Ramiro-Sousa N., Rodriguez-Martinez A., Sarri-Closa C., Sarró S., Larrubia, J., Andres-Bergareche H., Miguel-Cesma M., Padilla P., Salvador R., Pomarol-Clotet E., Hinzen W. [2019]: 'The Linguistic Signature of Hallucinated Voice Talk in Schizophrenia', *Schizophrenia Research* **206**, pp. 111–17.

Tsao, D. Y. and Livingstone, M. S. [2008]: 'Mechanisms of Face Perception', *Annual Review of Neuroscience*, **31**, pp. 411–437.

Vercammen, A., Knegtering H., den Boer J., Liemburg E., and Aleman A. [2010]: 'Auditory Hallucinations in Schizophrenia Are Associated with Reduced Functional Connectivity of the Temporo-Parietal Area', *Biological Psychiatry*, **67**, pp. 912–18.

Waters, F., Badcock J., Michie P., and Maybery M. [2006]: 'Auditory Hallucinations in Schizophrenia: Intrusive Thoughts and Forgotten Memories', *Cognitive Neuropsychiatry*, **11**, pp. 65–83.

Wilkinson, S. [2014]: 'Accounting for the Phenomenology and Varieties of Auditory Verbal Hallucination within a Predictive Processing Framework', *Consciousness and Cognition*, **30**, pp. 142–55.

Wilkinson, S. and Fernyhough C. [2017]: 'Auditory Verbal Hallucinations and Inner Speech: A Predictive Processing Perspective', in Radman Z. (*ed*.), *Before Consciousness: In Search of the Fundamentals of Mind.* Exeter: Imprint Academic, pp. 285–304.

Woods, A., Jones N., Alderson-Day B., Callard F., and Fernyhough C. [2015]: 'Experiences of Hearing Voices: Analysis of a Novel Phenomenological Survey', *The Lancet Psychiatry*, **2**, pp. 323–31.

Wu, W. [2012]: ‘Explaining Schizophrenia: Auditory Verbal Hallucination and Self-Monitoring’, *Mind & Language*, **27**, pp. 86–107.