

This is a repository copy of *From Computation to the First-Person : Auditory-Verbal Hallucinations and Delusions of Thought Interference in Schizophrenia-Spectrum Psychoses*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/185370/>

Version: Published Version

---

**Article:**

Humpston, Clara S, Adams, Rick A, Benrimoh, David et al. (10 more authors) (2019) From Computation to the First-Person : Auditory-Verbal Hallucinations and Delusions of Thought Interference in Schizophrenia-Spectrum Psychoses. *Schizophrenia bulletin*. S56-S66. ISSN 0586-7614

<https://doi.org/10.1093/schbul/sby073>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

## From Computation to the First-Person: Auditory-Verbal Hallucinations and Delusions of Thought Interference in Schizophrenia-Spectrum Psychoses

Clara S. Humpston<sup>\*1,2,3</sup>, Rick A. Adams<sup>3,4</sup>, David Benrimoh<sup>4,5</sup>, Matthew R. Broome<sup>6-8</sup>, Philip R. Corlett<sup>9</sup>, Philip Gerrans<sup>10</sup>, Guillermo Horga<sup>11</sup>, Thomas Parr<sup>4,9</sup>, Elizabeth Pienkos<sup>12</sup>, Albert R. Powers III<sup>9</sup>, Andrea Raballo<sup>13,14</sup>, Cherise Rosen<sup>15</sup>, and David E. J. Linden<sup>2,16</sup>

<sup>1</sup>Department of Psychological Medicine, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom; <sup>2</sup>School of Psychology, Cardiff University, Cardiff, United Kingdom; <sup>3</sup>Division of Psychiatry, University College London, London, United Kingdom; <sup>4</sup>Wellcome Trust Centre for Neuroimaging, University College London, London, United Kingdom; <sup>5</sup>Department of Psychiatry, McGill University, Montreal, Quebec, Canada; <sup>6</sup>Institute for Mental Health, School of Psychology, College of Life and Environmental Sciences, University of Birmingham, Birmingham, United Kingdom; <sup>7</sup>Department of Psychiatry, Medical Sciences Division, University of Oxford, Oxford, United Kingdom; <sup>8</sup>Faculty of Philosophy, Humanities Division, University of Oxford, Oxford, United Kingdom; <sup>9</sup>Yale School of Medicine, Yale University, New Haven, CT; <sup>10</sup>Department of Philosophy, The University of Adelaide, Adelaide, South Australia, Australia; <sup>11</sup>Department of Psychiatry, Columbia University, New York, NY; <sup>12</sup>Graduate Institute of Professional Psychology, University of Hartford, West Hartford, CT; <sup>13</sup>Department of Psychology, Faculty of Social and Educational Sciences, Norwegian University of Science and Technology, Trondheim, Norway; <sup>14</sup>Department of Medicine, Division of Psychiatry, Clinical Psychology and Rehabilitation, University of Perugia, Perugia, Italy; <sup>15</sup>Department of Psychiatry, University of Illinois at Chicago, Chicago, IL; <sup>16</sup>Division of Psychological Medicine and Clinical Neuroscience, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom

\*To whom correspondence should be addressed; Department of Psychological Medicine (PO72), Institute of Psychiatry, Psychology & Neuroscience, King's College London, Denmark Hill, London SE5 8AF, United Kingdom; tel: +44 (0) 20 7848 0088, fax: +44 (0) 20 7848 0298, e-mail: [clara.humpston@kcl.ac.uk](mailto:clara.humpston@kcl.ac.uk)

Schizophrenia-spectrum psychoses are highly complex and heterogeneous disorders that necessitate multiple lines of scientific inquiry and levels of explanation. In recent years, both computational and phenomenological approaches to the understanding of mental illness have received much interest, and significant progress has been made in both fields. However, there has been relatively little progress bridging investigations in these seemingly disparate fields. In this conceptual review and collaborative project from the 4th Meeting of the International Consortium on Hallucination Research, we aim to facilitate the beginning of such dialogue between fields and put forward the argument that computational psychiatry and phenomenology can in fact inform each other, rather than being viewed as isolated or even incompatible approaches. We begin with an overview of phenomenological observations on the interrelationships between auditory-verbal hallucinations (AVH) and delusional thoughts in general, before moving on to review several theoretical frameworks and empirical findings in the computational modeling of AVH. We then relate the computational models to the phenomenological accounts, with a special focus on AVH and delusions that involve the senses of agency and ownership of

thought (delusions of thought interference). Finally, we offer some tentative directions for future research, emphasizing the importance of a mutual understanding between separate lines of inquiry.

*Key words:* computational psychiatry/phenomenology/modeling/agency/ownership

*I think everyone hears a voice that's their thoughts. It's only when those voices start having their own mind and willpower that you're hearing voices in your head.*

(paranoid\_cataclysm, July 31, 2009; Raballo<sup>15</sup>)

### Introduction

Psychosis is a syndrome characterized by severe distortions in one's sense of reality. The most prominent symptoms of psychosis are delusions and hallucinations, which are usually defined as fixed false beliefs and perceptions without corresponding external stimuli, respectively. However, such definitions have been critiqued as rather arbitrary if

not oversimplistic and as failing to capture the full complexity and heterogeneity of the experiences of psychosis.

Both phenomenological and neuroscientific approaches have attempted, albeit in very different ways, to address some of this complexity in psychotic experiences. However, there is a substantial explanatory gap between these approaches.<sup>1</sup> One possibility for addressing this gap is computational psychiatry, a newly emerging field that uses formal mathematical models to delineate mechanisms of brain function and disease states.<sup>2</sup> This approach has the advantage of being able to bridge basic and clinical neuroscience research and potentially offer unifying accounts not only within a given psychiatric disorder but across different symptoms and diagnostic domains.<sup>3</sup> Here we begin by reviewing the relationships between auditory-verbal hallucinations (AVH) and delusional thoughts in schizophrenia-spectrum psychoses, before offering critical analyses of current theories from different levels of explanation.

We argue that delusional thoughts—in particular, delusions about the agency and ownership of thought (ie, delusions of thought interference)—and AVH are not best viewed as isolated mental events but are intricately related phenomena not only in degree but also in kind, ie, under certain circumstances delusions and hallucinations may not be clearly separable. This view was actually supported by French psychiatrist Esquirol,<sup>4</sup> who formally introduced the concept of hallucination in psychiatry in the 19th century and whose legacy has sadly been largely forgotten. Esquirol's notion was that hallucinations are a form of delirium and not a perception that “makes patients *believe* they have a perception (added italics),”<sup>5</sup> therefore stressing the belief-like and cognitive/intellectual aspects of hallucination. With Esquirol's original insight in mind, we aim to review computational models of AVH and evaluate how these models are best applied to phenomenological reports of AVH and delusions of thought interference (eg, delusions resulting from an experience of thought insertion), before arguing that these symptoms may just be different manifestations of intrinsically similar neural processes, eg, processes using Bayesian probabilistic inference.<sup>6</sup> For more details about the background, see [Supplementary Materials](#).

### The Blurred Lines Between Thought and Perception

Although they are usually allocated into separate descriptive silos (ie, according to the conventional dichotomy between aberrant *perception* and *cognition*, respectively), AVH and delusional thoughts share important phenomenological features, such as an autocentric, self-referential architecture and profound alterations of lived space, time, and intersubjectivity.<sup>7–12</sup> This is particularly manifest in prototypical “voices” (eg, commenting and imperative voices) and transitive delusions or thought interference (eg, delusions of control, thought insertion, thought withdrawal, and thought broadcasting), and also in delusions of reference and persecution. Recent empirical

analyses<sup>13</sup> indicate indeed that AVH articulate themselves in an experiential realm that is in-between the phenomenology of cognition and perception, retaining features of both an altered stream of thought and quasi-material aspects of sensorial givenness.<sup>12,14–16,61</sup>

The psychopathological interconnectedness of delusional thoughts and AVH may be somewhat circular (at least in the sense of potential co-perpetuation): the former may increase the proclivity to thematize anomalies of the stream of consciousness as AVH, and, conversely, AVH could promote the further articulation of delusional themes. Moreover, any account that posits delusional thoughts and AVH as descriptively separable symptoms needs to take heed of autobiographic accounts<sup>18</sup> as well as phenomenological research,<sup>19,20</sup> suggesting that both are consequences of a psychotic transformation of the medium of consciousness. This was relatively clear for major authors of the last century; eg, Bleuler<sup>21</sup> emphasized that AVH operate through a comprehensive transitive delusional ideation: “the voices not only speak to the patient, but they pass electricity through his body, beat him, paralyze him, take his thoughts away” (1911/50, p.94). Minkowski<sup>22</sup> characterizes schizophrenia as arising from the *trouble générateur* and a “loss of vital contact with reality.” Even Jaspers,<sup>23</sup> who suggested primary schizophrenic delusions may be outside the realm of “understandability” (although secondary delusions may still be amenable to empathic understanding), insisted that the task of psychopathology was to focus on “actual conscious psychic events,” rather than isolated, clinician-identified symptoms no matter how un-understandable they may seem to be. Yet, contemporary empirical research has rediscovered such intersection between AVH and delusional thoughts only recently.<sup>13,24</sup>

One prominent contemporary theory of schizophrenia that incorporates this perspective was developed by Sass and colleagues,<sup>25,26</sup> who view symptoms of schizophrenia as manifestations of a disturbance of *ipseity* or basic self-experience, ie, of being a “vital and self-coinciding *subject* of experience” (p. 428). Thought, perception, and action come to feel strange, awkward, foreign, extrinsic to, and alienated from oneself.<sup>27,28</sup> Fuchs<sup>29</sup> has put forward a somewhat different hypothesis, arguing that both AVH and delusional thoughts arise from a transformation of the world in which worldly objects lose their independent existence (ie, they are no longer available to others and perceivable from a variety of other perspectives), which can result in objects seeming to exist only for oneself, as well as the experience of special self-directed meanings or messages associated with objects and events (delusions). In this context, perception-like phenomena that are not in fact accessible to others (hallucinations) may take on an ontological status similar to what others may consider to be “objective” perceptions. (However, as Ratcliffe<sup>30</sup> notes, these accounts do not address some potentially crucial aspects of hallucinations, including their specific (and often quite negative) content, the perceptual (or quasi-perceptual) qualities of hallucinations

(compared to other manifestations of disturbed ipseity), and, perhaps most importantly, the fact that hallucinations occur not only for persons with ipseity disorders (ie, schizophrenia), but also for persons with trauma histories, mood disorders, or no clinical history at all.) It should be noted that hallucinations, at least when described in schizophrenic psychoses, often retain centrality of the self or a kind of solipsistic and subjectivized quality<sup>31–34</sup> and, for many, are distinguishable from typical perceptions.<sup>35,36</sup>

Indeed, it has been argued that how patients with schizophrenia access conscious information (including an elevated threshold to consciousness) plays a crucial role in the merging of cognition and perception, and it is likely that a unifying mechanism underpins both processes. The proposal of Northoff and Huang<sup>37</sup> of a temporospatial theory of consciousness (TTC) links the 4 dimensions of consciousness (level/state, content/form, phenomenology/experience, and cognitive processing/reporting) and offers an account of the brain as situated in time and space while constructing its own temporospatial structure. In particular, the concept of “temporospatial *alignment*” is considered disrupted in schizophrenia relating to the content and form of consciousness.<sup>38</sup> Although the TTC does not explicitly incorporate the relationships between consciousness and self, the brain regions (including cortical midline structures) thought to underlie self-experience are also at least partly responsible for the alignment and integration of conscious experience. This offers tentative evidence that there may be a common basis underlying thought and perception that are embedded in the very structure of consciousness. As Henriksen et al<sup>27</sup> note, “From a phenomenological perspective, AVHs in schizophrenia are not primarily sensory-perceptual but rather cognitive phenomena arising from a partial dissolution of certain structures of self-consciousness”(p.166).

Such work may be used to frame and guide empirical research on schizophrenia and hallucinations; thus, computational models of hallucinations and other psychotic symptoms should be able to account for and explain not only isolated symptoms but also symptoms as they are experienced, as interrelated and embedded within an overall context of patients’ selfhood and relationship with the world. As Larøi et al<sup>39</sup> note, future work on hallucinations must acknowledge their status as “meaningfully interrelated facets of a more comprehensive and characteristic gestalt change in the patient’s experience (field of consciousness) and existence” (p. 235). In the section “Modeling AVH: Computational and Cognitive Frameworks,” we review some of the most prominent models in the computational psychiatry of AVH research and consider how they may fit in the wider framework of altered subjective experiences.

### Modeling AVH: Computational and Cognitive Frameworks

Any model of hallucinations must be constrained by empirical neurobiological findings, such as those reviewed

in [Supplementary Materials](#). In this section, we first provide a brief overview of a network model of AVH, followed by one of the most dominant cognitive models of AVH (the inner speech model) while integrating it with neuroscientific and computational theories. Finally, we focus on a relatively recent hierarchical Bayesian model of AVH and how it relates to our understanding of AVH.

#### *Attractor/Network Models*

An attractor is a set of configurations to which the states of a dynamical system are drawn.<sup>40</sup> Examples in neuroscience include the attractors used to model working memory processes in the prefrontal cortex<sup>17,41–43</sup> and oculomotor control<sup>44</sup> and to account for the activities of hippocampal cells involved in navigation.<sup>45,46</sup> They have also been used to model features of schizophrenia.<sup>47,48</sup>

The dynamics of a system can be described in terms of its “energy” landscape. Attractors manifest as regions of low energy (basins) in the energy landscape. In a neural network, the shape of the landscape is determined by the strength of connectivity between neuronal populations. This concept complements many other modeling approaches. For instance, theories founded on Bayesian inference either implicitly or explicitly rely on the existence of minima in an energy functional, specifically, a free energy (cf predictive coding<sup>49</sup>) or an approximation of this (cf loopy belief propagation<sup>50</sup>).

Perceptual inference can be formulated as a process of descent from high to low energy.<sup>51–53</sup> False perceptions, such as hallucinations, might in theory result from changes in the shape of the energy landscape. It has been proposed that disruptions in GABA and NMDA receptor conductance (modulated by dopamine<sup>54</sup>) in the prefrontal cortex could result in such changes.<sup>55</sup> For example, a reduction of excitatory pyramidal-inhibitory interneuron connectivity would mean that other states are less suppressed when one group of neurons is active. An increase in stochastic fluctuations in firing rates would make it easier to jump from one basin to another.

This proposition potentially accounts for several phenomena in schizophrenia including hallucinations, while also accommodating prominent neurobiological theories of disrupted excitation-inhibition (E-I) balance<sup>56,57</sup> and aberrant dopamine signaling.<sup>58</sup> An alternative to modulating synapses to disrupt the energy landscape is to remove connections between neurons. Such “pruning” approaches have successfully reproduced (single word) hallucinations *in silico*, but neither model has so far demonstrated the spontaneous production of more fluent speech.<sup>59</sup>

#### *Inner Speech/Comparator Model*

There are good reasons to think that AVH involve motor processes. Perhaps the most convincing is that they involve the consequences of an action (ie, speech), in

contrast to other forms of hallucinations. This is another reason for our focus on AVH, as nonverbal hallucinations are very likely to have different mechanisms.<sup>60,61</sup> The “comparator” or “inner speech” model is an influential model of AVH based on predictive motor processes—efference copy and the forward model. It proposes that efference copies of motor commands are sent to a “forward model” that uses them to predict their sensory consequences in advance of sensory feedback. Successful prediction may attenuate the perception of those sensory consequences and also result in the feeling of agency for a movement, whereas prediction errors may lead one to infer the environment or some other agent was responsible for the discrepancy. Patients with schizophrenia are known to have problems with prediction (eg, of the motion of targets during smooth pursuit),<sup>62</sup> and if these problems extend to predicting the consequences of one’s own movements, then the consequent loss of the feeling of agency for self-generated movements could lead to the belief that another agent is responsible for them: ie, passivity symptoms.<sup>63</sup> In terms of AVH, the sensation that AVH come from someone else/another source other than one’s self could also be accounted for by failures in self-monitoring. The externality and alien nature of AVH are indeed important phenomenological features in many—but not all—cases. This model is bolstered by numerous empirical findings.<sup>64,65</sup> In addition, the self- and source-monitoring models of AVH have benefited from studies using signal detection theories and emotional processing, where a sense of perceptual hypervigilance linked to heightened emotional states (threat, fear, and anxiety) results in an urgent need to reduce uncertainty of the signal and lowers the threshold of auditory perception, especially when it comes to internally generated stimuli.<sup>66</sup> Interestingly, this kind of “jumping-to-conclusions” or rapid judgments based on limited sensory or cognitive evidence are also key to delusion formation.<sup>67</sup>

The ipseity disturbance model suggests that such source-monitoring disturbances could be a reflection of the hyperreflexivity or tendency to take tacit acts of consciousness as an object of reflection rather than the implicit medium of awareness of the world<sup>25</sup> (although Henriksen et al<sup>27</sup> contest the assumption in source-monitoring theories that patients perceive their hallucinations as real). This model points to a notion of selfhood that it maintains a world- or object-directed intentionality; when intentionality is no longer prereflectively inhabited, thoughts or other intentional acts may become distorted and take on physical or perceptual qualities. Indeed, early experiences of subtle changes in the experience of cognition and stream of consciousness, including difficulties distinguishing between thought and perception, may predict the later development of hallucinations.<sup>68</sup> Other models suggest that AVH may be more related to a disruption of anticipation of certain kinds of anxiety-producing thoughts: certain unwanted thoughts may be

anxiously anticipated and therefore imbued with unusual object- and perception-like qualities.<sup>69</sup> This model may help to explain the content-specific nature of many AVH, suggesting that it is not all thoughts that feel alienated and external, but only those with particularly disturbing or distressing content. An alternative model may be a better explanation for false inference about communications from another agent; thus, we next discuss the hierarchical Bayesian model of AVH. For some of the critiques on the comparator model, see [Supplementary Materials](#).

### *A Hierarchical Bayesian Model of AVH*

A popular view of the brain among contemporary neuroscientists is that it instantiates a hierarchical Bayesian model of its environment. This view has several important implications for hallucination research.<sup>70</sup> First, different levels in the brain’s hierarchical organization represent the causes of sensory data at levels below in an increasingly abstract way as one ascends the hierarchy.<sup>71</sup> Second, the brain uses or approximates Bayesian inference to infer these causes: meaning it must combine prior beliefs about these causes with sensory data (in the form of a “likelihood”)—weighted by their relative certainty (or “precision”)—to make its inferences, or “posterior beliefs.” Crucially, the incorrect assignment of precision leads to failures of inference, eg, if sensory precision is underestimated or prior precision overestimated, then the prior will dominate the posterior: a potential cause of hallucinations of any sort.<sup>49</sup> Third, most priors within the model—eg, the expected sound of someone’s voice or the content of their speech—are learned from previous inferences: known as “empirical priors.”

Numerous neural message-passing schemes can perform Bayesian inference. One such is predictive coding, in which descending messages from higher levels are predictions of quantities at lower levels, and ascending messages are prediction errors—the difference between the predicted and the actual values—weighted by their precision.<sup>49</sup> An alternative is belief propagation, in which the descending and ascending messages are priors and likelihoods.<sup>72</sup> If the brain uses predictive coding, how might it encode precision? Given precision changes the weights of priors and likelihoods in inference without altering their means, its neural implementation ought to increase the “gain” of neural messages without generating new messages itself. One obvious candidate mechanism is synaptic gain, ie, the factor by which presynaptic input is multiplied to generate postsynaptic potentials. Synaptic gain can influence the precision of encoded states at different scales: both at the neural level, via neuromodulatory receptors such as the *N*-methyl-D-aspartate (NMDA) receptor and receptors for dopamine and acetylcholine, and at the network level, by altering the robustness of neural representations to other inputs or stochastic

fluctuations in neural activity. In particular, NMDA receptors on inhibitory interneurons—thought to be dysfunctional in schizophrenia<sup>73</sup>—may determine the extent to which one neural “explanation” for sensory input can suppress competing “explanations.”

Many paradigms—eg, visual illusions, oculomotor pursuit tasks, EEG oddball tasks, sensory attenuation tasks, and belief updating tasks—indicate that patients with schizophrenia have an imbalance between the precision of priors (too low) and sensory evidence (too high).<sup>74</sup> In other paradigms in schizophrenia, however, priors dominate sensory evidence<sup>75,76</sup>—interestingly, in both cases these were recently learned perceptual priors. There are various possible explanations for these apparently opposite imbalances in different paradigms. One is that the hierarchy is so deep that, eg, decreases in midlevel synaptic gain could be expressed as decreased precisions of either priors (for the levels below, which receive predictions from this level) or likelihoods (for the levels above, which receive prediction errors from this level), depending on whether the paradigm depends more on lower or higher levels for its effects. Another is that schizophrenia may involve a failure of adaptive *control* of synaptic gain, rather than simply too much or too little at different levels. Belief propagation account of Jardri and colleagues proposes these imbalances come about due to priors or likelihoods being “overcounted” due to disinhibited message passing,<sup>76–78</sup> rather than alterations in synaptic gain, although this and other belief propagation models also contain precision terms that could be encoded by synaptic gain.

Under a hierarchical Bayesian account of AVH, hallucinations must result from overprecise priors that are not corrected by (relatively imprecise) likelihoods (ie, the precision of the mapping between auditory input and its possible causes is low).<sup>71</sup> This is because a false-positive inference requires that a percept is internally generated and is not derived from ascending sensory information. An interesting question is where such priors might be: Are they at the higher levels of the model (eg, prefrontal cortex), imposing themselves on middle levels, or are they at middle levels (eg, superior temporal cortex), imposing themselves on lower levels (ie, empirical priors)? The latter could be an instance of failure to attenuate sensory precision (ie, synaptic gain), as is found in the force-matching paradigm.<sup>64,65,79</sup> In this case, increased precision/gain may be of autonomous neural activity in higher auditory cortex (as proposed by “attractor state” models reviewed in “Attractor/Network Models” section) rather than of sensory input. The former case may result from a compensatory increase in prior precision.<sup>80</sup> Recent modeling of the former option using a Bayesian framework indicates that AVH can occur when an agent engaging in dialogue has a strong prior belief that it is listening to a voice and an imprecise likelihood mapping. When it expects to hear a voice but none is present, it generates a hallucinated voice

to satisfy its expectations.<sup>81</sup> The fact that this expected voice is that of a conversational partner may also play a role in explaining aberrancies of agency in AVH.<sup>81</sup>

Numerous kinds of priors could contribute to AVH. A large cluster analysis of AVH phenomenology (largely schizophrenia) indicates different subtypes exist—memories, first-person AVH, other-person AVH, and unintelligible sounds<sup>82</sup>—even within one person. This suggests that the common pathology lies in an area that receives predictions about speech from multiple sources (ie, superior temporal cortex) but is unable to sufficiently attenuate their precision, such that these priors dominate resulting percepts. This will especially be the case if the precision of incoming sensory data is low. In the case of memories and first-person AVH, predictions are likely to come from medial temporal lobe and a network of frontal areas including inferior frontal gyrus<sup>40,83</sup> (also associated with “inner speech”), respectively.

Given that AVH develop increasing complexity over time,<sup>84</sup> the likeliest explanation is that identities of other-person voices are empirical priors that are inferred from previous hallucinations. For example, hearing a muffled insult might create an image of someone whose voice that might be, which may generate expectations about other insults. The development of empirical priors concerning the origins of AVH will have a critical impact on the extent to which the hallucinating subject feels he or she has control over the AVH—a key distinguishing factor between psychotic and healthy voice-hearers.<sup>84,85</sup> If voices that could be ascribed to distorted perceptions or hallucinations are instead attributed to an external agent, the voice-hearer’s sense of autonomy will diminish.

Another key phenomenological difference is the emotionally negative content (usually) associated with clinical or psychotic voice-hearers; although this has not yet been systematically studied under the Bayesian framework, there is accumulating evidence pointing toward the roles played by past memories (eg, of trauma), which may act as top-down expectations relating to voice content,<sup>80,86</sup> especially in memory-related subtypes of AVH.

This model makes predictions that cohere with empirical investigations: (1) Loss of NMDA receptor function on inhibitory interneurons in temporal areas might lead to an increase in resting activity (it is harder to suppress “noise”) but a decrease in task-related (including prediction error-related<sup>86</sup>) activity (task-related activity is inversely correlated with the baseline level)—both of which have been demonstrated in functional magnetic resonance imaging studies with schizophrenia patients with AVH,<sup>86</sup> but are hard to explain using the “comparator” model.<sup>80</sup> (2) Inhibitory interneuron dysfunction in schizophrenia may make it harder for auditory cortex to generate  $\gamma$  oscillations in response to 40-Hz tones (which depend on interactions between pyramidal cells

and interneurons).<sup>87</sup> (3) Voice-hearers ought to demonstrate more precise empirical priors than controls—this has been shown in both hallucinating patients with schizophrenia and healthy voice-hearers, using tasks in which learned contexts influence uncertain sensory data<sup>88</sup> (see figure 1; also Cassidy et al<sup>89</sup>).

The Bayesian model of AVH may have phenomenological correlates in the notion of aberrant salience, which reflects a tendency to attend to irrelevant or background perceptual details (as opposed to details that are more relevant for goal-oriented behavior). It has been suggested that persons with schizophrenia may demonstrate this anomalous processing of perceptual stimuli. Such attentional and perceptual disturbances may result in the assignment of greater importance to irrelevant stimuli,<sup>90</sup> rather than allowing one's attention to be directed (and corrected) by more accurate information and prior knowledge about the environment. The ipseity disturbance model suggests that this salience disruption results from a loss of “grip” or “hold” on one's engagement with the environment, whereby disturbances in basic selfhood are likely to disrupt one's pragmatic and goal-oriented engagement with the world.<sup>26</sup> It may also impact the inability to appropriately correct or update one's interpretations about various perceptual inputs, and the tendency to interpret ambiguous or “noisy” stimuli according to strong overprecise (perhaps emotionally salient) priors.<sup>69</sup>

### AVH and Delusions of Thought Interference

Whether a thought has vocal, audible, or agentic qualities may partly account for the content of the delusion and the way it is reported in research protocols and clinical encounters. Typically, audible voices will be described as AVH and soundless voices as thought insertion.<sup>12</sup> Yet such descriptions may impose an artificially precise structure on an intrinsically ambiguous form of experience, which also includes emotional distress and the elaborations, interpretations, and defenses produced by the patient. The above theories are compelling in their ability to explain the development and maintenance of AVH from the perspectives of neurobiology, perceptual and motor systems, and cognitive modeling. In this section, we consider how some of these models might be able to shed light on several major phenomenological features of AVH and thought insertion, especially findings of their interrelationship and experiential similarities.

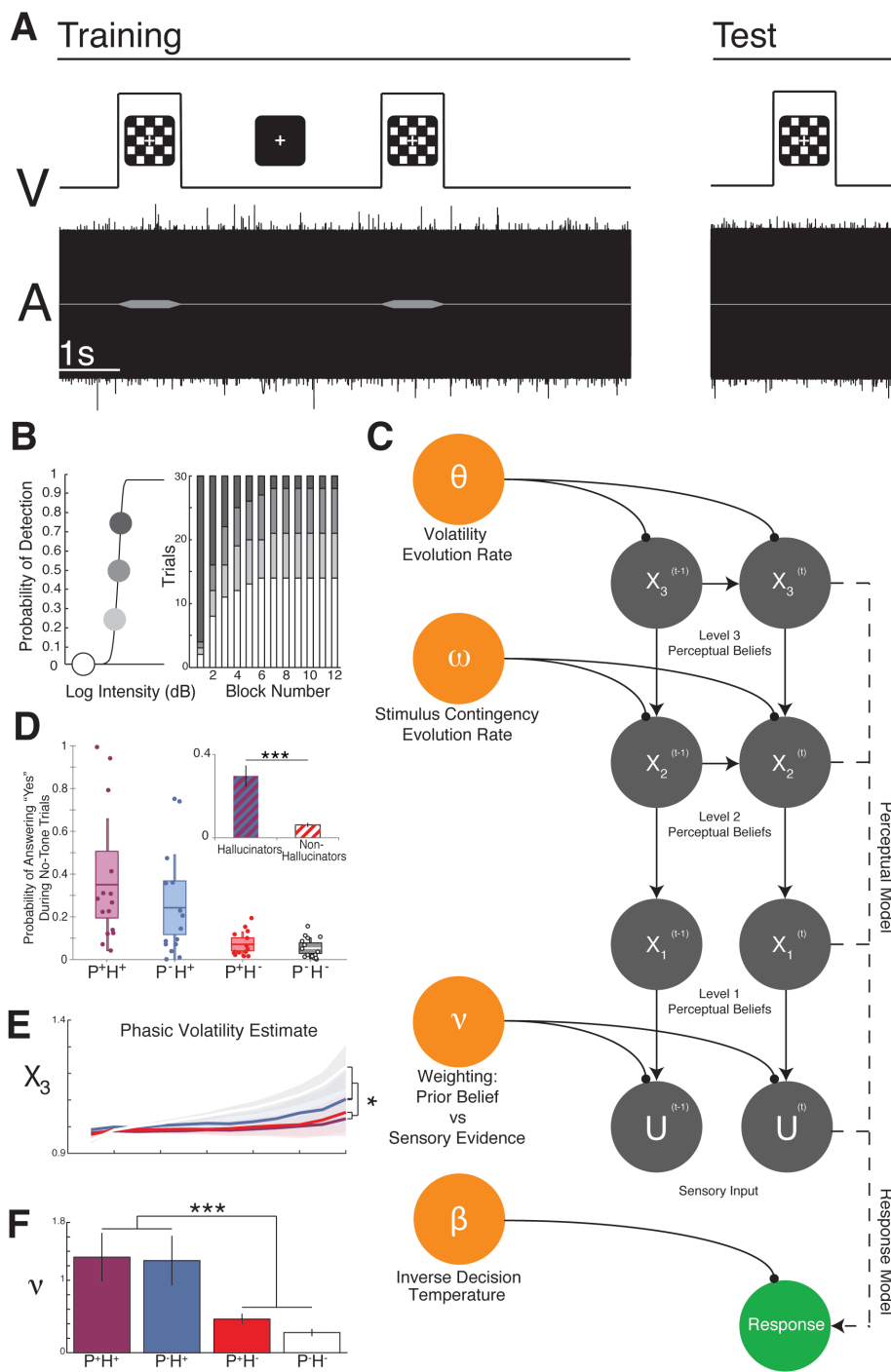
In thought insertion, the subject reports that thoughts arrive in their mind “out of nowhere.”<sup>91</sup> Most explanations of thought insertion descend from source-monitoring accounts of intentional action.<sup>92,93</sup> These accounts distinguish sense of ownership (the experience that the action belongs to oneself, that the bodily movement is one's own) from sense of agency (the experience of intending and controlling the action).<sup>94,95</sup> When sense of agency is absent, but sense of ownership is intact,

the patients experience their body performing actions, which are nonetheless not felt to be theirs. This way of conceptualizing the experience treats sense or ownership as a form of subjective awareness of bodily occupation of space.<sup>96</sup> It also is worth noting that some authors<sup>94</sup> propose the sense of agency and ownership are in fact 2-fold, consisting of a “feeling” and a “judgment” that incorporate bottom-up sensory information (feeling) and top-down beliefs (judgment). From this perspective, one could have a mere feeling of intending the movement but deny the movement is indeed initiated by themselves (due to higher-order factors or conflicts perhaps), such as in experiences associated with passivity phenomena. Here, we use the term “sense” to refer to the prereflective *feeling* of agency/ownership in Synofzik's account.

Source-monitoring accounts of thought insertion decompose the experience of subjectivity into a sense of ownership and a sense of agency for *thought* (SOT and SOAT, respectively). Thought insertion, on this view, derives from the experience of SOT for thought unaccompanied by SOAT.<sup>97</sup> There is, however, no straightforward analogy between thoughts and actions. Actions are monitored, at different levels of cognition, for consistency with intentions, generating (in this account) a sense of agency. But we do not, in general, intend our thoughts. So SOAT cannot depend on comparison of actual and intended/predicted thoughts.<sup>98</sup>

Nonetheless, the analogy has been pursued in different ways, each of which argues that a characteristic SOAT goes missing in thought insertion, leaving SOT for thoughts intact. The source of a SOAT might be a sense of agency for speech production (on the plausible assumption that the thoughts in question are episodes of inner speech),<sup>99,100</sup> protention (online anticipation of the temporal structure of episodes of thought)<sup>101</sup> or egotonia (compatibility with self-representation).<sup>102</sup> The suggestion is that a monitoring process fails to detect a match between an occurrent thought and the relevant generative process, as outlined in the “Inner Speech/Comparator Model” section.<sup>103</sup> Another proposal is that source monitoring is between background psychology (understood as the totality of tacit representations on which thinking depends) and conscious thought.<sup>96</sup>

A difficulty with all these proposals is accounting for normal and pathological (eg, in obsessive compulsive disorder) cases of unbidden or intrusive thoughts. Such thoughts arise in the mind in an unpredicted manner but are not experienced or explained as inserted. Another difficulty with such proposals is more fundamental. To preserve the Cartesian intuition, their proponents divide subjectivity into SOAT and SOT, arguing that SOT is intact in thought insertion. It is not clear that this reflects patients' experience. Patients say that the thoughts are “not theirs.”<sup>91,97</sup> If we take this seriously, then perhaps the idea of an intact SOT with its connotation of intact “inner space” or mental boundaries needs to be rethought. One



**Fig. 1.** A hierarchical Bayesian model of conditioned hallucinations. Reproduced with permission. In this study,<sup>88</sup> subjects with and without psychosis ( $P^+$ ,  $P^-$ , respectively) and with and without auditory verbal hallucinations ( $H^+$ ,  $H^-$ , respectively) were conditioned to associate tones presented at threshold with a concurrent checkerboard stimulus (A). This association was then tested by presenting the checkerboard alone and recording subjects' reports of hearing the tone: the probability of presentation of a subthreshold or absent tone increased over blocks (B). Subjects with hallucinations (irrespective of psychosis) were more likely to hallucinate tones (D,  $***P < .001$ ). The evolution of this association over time was modeled using a hierarchical Bayesian model (C), in which the first level ( $X_1$ ) is the belief in the presence of the tone, the second level ( $X_2$ ) is the association of the tone with the checkerboard, and the third ( $X_3$ ) is the rate of change of the association, and subject-specific parameters  $\omega$  and  $\theta$  affect how quickly these beliefs change at levels 2 and 3. At  $X_3$ , there was a significant block-by-psychosis interaction (E,  $*p < 0.05$ ). The belief about having heard a tone (ie, at  $X_1$ ) on one trial—in Bayesian terms, the “posterior”—then becomes the prior belief in the next trial, and the weighting of this prior belief over the incoming sensory evidence is parameterized by  $\nu$ . Crucially,  $\nu$  was higher in hallucinators (irrespective of psychosis): ie, hallucinators overweighted their (empirically learned) prior beliefs in this task (F,  $***P < .001$ ). One could likewise model voice perception using a hierarchical model in which inputs are sounds and higher levels encode beliefs about phonemes, words, sentences, and speaker identities. In such a model, verbal hallucinations could result from a similar overweighting of prior beliefs.<sup>104</sup>



possibility, consistent with recent Bayesian approaches to cognition, is that the boundary of inner space is not an intrinsic or given feature of cognition but is actively constructed by the mind. Henriksen et al<sup>27</sup> suggest that persons diagnosed with schizophrenia experience their interiority as an unusually concrete form of inner space, possibly due to disruption of a natural, embodied relation to exterior space, with the result that thoughts are given spatial, object-like qualities. Similarly, whether a representation is experienced as internal or external to the mind may depend on the predictive model deployed for the task; persons diagnosed with schizophrenia may inappropriately use external models for internal events.

Some ingenious paradigms addressed this issue by defining neural correlates of AVH and thought insertion. Jardri et al<sup>101</sup> found that transcranial magnetic stimulation over the left temporal–parietal junction (TPJ) reduced AVH and the right TPJ improved agency for inner speech. This result is consistent with others suggesting that sense of agency depends on modulation of inferior parietal areas by TPJ. Note however that the subjects here did not report thought insertion. The clearest result in this area is from an experiment contrasting hypnotically induced thought insertion, which found “specifically, reduced activity in language production regions, and *not* over-activation of cerebellar-parietal regions, was present during thought insertion” (added italics).<sup>100</sup> The same experiment found reduced activation in supplementary motor area in experience of both thought insertion and delusions of alien control.

Interestingly, phenomenological theories of psychosis provide little clarification regarding similarities and differences between AVH and thought insertion. One suggestion is that there is little or no meaningful difference between these two phenomena, or at least between thought insertion and the more thought-like, inner speech forms of AVH.<sup>69,105</sup> Instead, it is suggested that the different ways they are represented or described by patients may simply reflect individual or cultural differences in the way anomalous experiences are conceptualized and communicated.<sup>105,106</sup> However, advances in psychosis research including those on neural computation might be useful for clarifying the shared (or disparate) phenomenology of thought insertion and various forms of AVH.

## Conclusion

Some tentative suggestions for further research and limitations of the current approach can be found in [Supplementary Materials](#). In this conceptual review, we present some of the phenomenological observations and computational models in the study of AVH and delusions of thought interference in psychosis. These have a particular focus on the merging between (delusional) thought and (hallucinatory) perception, which are embedded in current theories of both computational psychiatry and

phenomenological psychopathology. We argue that these two approaches are indispensable to advancing the study of psychosis and urge researchers and clinicians to keep the patients’ reality in mind when considering models and explanations. We are fully aware that incorporating different lines of inquiry is only the beginning of a dialogic process between seemingly disparate fields in psychopathology research, and there is no single approach that can account for the sheer complexity of mental illnesses such as schizophrenia. Nevertheless, interdisciplinary research and integrative approaches are crucial in creating a mutual understanding and facilitating collaborative efforts, with the ultimate goal of maximizing benefits to the quality of life of patients with psychosis. The realization that phenomenology and computation work in conjunction and inform each other is undoubtedly a vital first step toward achieving this goal.

## Supplementary Material

Supplementary material is available at *Schizophrenia Bulletin* online.

## Funding

This work was supported by a Medical Research Council (UK) Doctoral Training Grant (MR/K5013471/1) allocated to C.S.H.; D.E.J.L. is supported by the MRC Centre for Neuropsychiatric Genetics and Genomics (G0800509), Cardiff University; R.A.A. is funded by the Academy of Medical Sciences (AMS-SGCL13-Adams) and the National Institute of Health Research (CL-2013-18-003); T.P. is funded by the Rosetrees Trust (173346); D.B. is supported by the Richard and Edith Strauss Foundation and the Fonds de Recherche du Québec- Santé.

## Acknowledgments

The authors have declared that there are no conflicts of interest in relation to the subject of this study.

## References

1. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci*. 2012;16:72–80.
2. Friston KJ, Stephan KE, Montague R, Dolan RJ. Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry*. 2014;1:148–158.
3. Waters F, Fernyhough C. Hallucinations: a systematic review of points of similarity and difference across diagnostic classes. *Schizophr Bull*. 2017;43:32–43.
4. Esquirol E. *Mental Maladies: A Treatise on Insanity*. Philadelphia, PA: Lea and Blanchard; 1845.
5. Telles-Correia D, Moreira AL, Gonçalves JS. Hallucinations and related concepts-their conceptual background. *Front Psychol*. 2015;6:991.

6. Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci*. 2009;10:48–58.
7. Stanghellini G, Ballerini M, Blasi S, et al. The bodily self: a qualitative study of abnormal bodily phenomena in persons with schizophrenia. *Compr Psychiatry*. 2014;55:1703–1711.
8. Stanghellini G, Ballerini M, Presenza S, et al. Psychopathology of lived time: abnormal time experience in persons with schizophrenia. *Schizophr Bull*. 2016;42:45–55.
9. Northoff G, Stanghellini G. How to link brain and experience? Spatiotemporal psychopathology of the lived body. *Front Hum Neurosci*. 2016;10:76.
10. Henriksen MG, Nordgaard J. Self-disorders in schizophrenia. In: Stanghellini G, Aragano M, eds. *An Experiential Approach to Psychopathology*. Switzerland: Springer; 2016:265–280.
11. Handest P, Klimpke C, Raballo A, Larøi F. From thoughts to voices: understanding the development of auditory hallucinations in schizophrenia. *Rev Philos Psychol*. 2016;7:595–610.
12. Humpston CS, Broome MR. The spectra of soundless voices and audible thoughts: towards an integrative model of auditory verbal hallucinations and thought insertion. *Rev Philos Psychol*. 2016;7:611–629.
13. Woods A, Jones N, Alderson-Day B, Callard F, Fernyhough C. Experiences of hearing voices: analysis of a novel phenomenological survey. *Lancet Psychiatry*. 2015;2:323–331.
14. Raballo A, Larøi F. Murmurs of thought: phenomenology of hallucinatory consciousness in impending psychosis. *Psychosis*. 2011;3:163–166.
15. Raballo A. The stream of hallucinatory consciousness: when thoughts become like voices. *J Conscious Stud*. 2016;23:132–143.
16. Raballo A. From perception to thought: a phenomenological approach to hallucinatory experience. *Schizophr Bull*. 2017;43:18–20.
17. Strogatz SH. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*. Boulder, CO: Westview Press; 2014.
18. Saks E. *The Centre Cannot Hold: A Memoir of My Schizophrenia*. London, UK: Virago; 2007.
19. Schneider K. *Clinical Psychopathology*. New York: Grune & Stratton; 1959.
20. Ey H, Bernard P, Brisset C. Manuel de psychiatrie. *Rev Philos France Let*. 1972;162:458–459.
21. Bleuler, E. (1950). *Dementia praecox or the group of schizophrenias*. Oxford, UK: International Universities Press.
22. Broome MR. *The Maudsley Reader in Phenomenological Psychiatry*. Cambridge, UK: Cambridge University Press; 2013.
23. Jaspers K. *General Psychopathology*. Chicago, IL: University of Chicago Press; 1963.
24. Rosen C, Jones N, Chase KA, Gin H, Grossman LS, Sharma RP. The intrasubjectivity of self, voices and delusions: a phenomenological analysis. *Psychosis*. 2016;8:357–368.
25. Sass LA, Parnas J. Schizophrenia, consciousness, and the self. *Schizophr Bull*. 2003;29:427–444.
26. Sass L, Borda JP, Madeira L, Pienkos E, Nelson B. Varieties of self disorder: a bio-pheno-social model of schizophrenia. *Schizophr Bull*. 2018;44:720–727.
27. Henriksen MG, Raballo A, Parnas J. The pathogenesis of auditory verbal hallucinations in schizophrenia: a clinical–phenomenological account. *Philos Psychiatr Psychol*. 2015;22:165–181.
28. Ratcliffe M, Broome M. Existential Phenomenology, Psychiatric Illness and the Death of Possibilities. In: Crowell S, ed. *The Cambridge Companion to Existentialism*. Cambridge Companions to Philosophy. Cambridge, UK: Cambridge University Press; 2012:361–382.
29. Fuchs T. The intersubjectivity of delusions. *World Psychiatry*. 2015;14:178–179.
30. Ratcliffe, M. *Real hallucinations: psychiatric illness, intentionality, and the interpersonal world*. Cambridge, MA: MIT Press; 2017.
31. Sass LA. *The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind*. Ithaca, NY: Cornell University Press; 1994.
32. Parnas J, Sass LA. Self, solipsism, and schizophrenic delusions. *Philos Psychiatr Psychol*. 2001;8:101–120.
33. Parnas J, Henriksen MG. Mysticism and schizophrenia: a phenomenological exploration of the structure of consciousness in the schizophrenia spectrum disorders. *Conscious Cogn*. 2016;43:75–88.
34. Humpston CS. The paradoxical self: awareness, solipsism and first-rank symptoms in schizophrenia. *Philos Psychol*. 2018;31:210–231.
35. van den Berg JH. On hallucinating: critical-historical overview and guidelines for further study. In: de Koning AJJ, Jenner FA, eds. *Phenomenology and Psychiatry*. London: Academic Press; 1982:97–110.
36. Straus EW. Aesthesiology and hallucinations. In: May R, Angel E, Ellenberger HF, eds. *Existence*. New York: Simon & Schuster; 1958:139–169.
37. Northoff G, Huang Z. How do the brain's time and space mediate consciousness and its different dimensions? Temporospacial theory of consciousness (TTC). *Neurosci Biobehav Rev*. 2017;80:630–645.
38. Lakatos P, Schroeder CE, Leitman DI, Javitt DC. Predictive suppression of cortical excitability and its deficit in schizophrenia. *J Neurosci*. 2013;33:11692–11702.
39. Larøi F, de Haan S, Jones S, Raballo A. Auditory verbal hallucinations: dialoguing between the cognitive sciences and phenomenology. *Phenomenol Cogn Sci*. 2010;9:225–240.
40. Raij TT, Valkonen-Korhonen M, Holi M, Therman S, Lehtonen J, Hari R. Reality of auditory verbal hallucinations. *Brain*. 2009;132:2994–3001.
41. Braver TS, Barch DM, Cohen JD. Cognition and control in schizophrenia: a computational model of dopamine and prefrontal function. *Biol Psychiatry*. 1999;46:312–328.
42. Wimmer K, Nykamp DQ, Constantinidis C, Compte A. Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nat Neurosci*. 2014;17:431–439.
43. Wang XJ. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci*. 2001;24:455–463.
44. Seung SH. Continuous attractors and oculomotor control. *Neural Networks*. 1998;11:1253–1258.
45. Zhang K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J Neurosci*. 1996;16:2112–2126.
46. Wills TJ, Lever C, Cacucci F, Burgess N, O'Keefe J. Attractor dynamics in the hippocampal representation of the local environment. *Science*. 2005;308:873–876.
47. Hoffman RE, McGlashan TH. Neural network models of schizophrenia. *Neuroscientist*. 2001;7:441–454.

48. Rolls ET, Loh M, Deco G, Winterer G. Computational models of schizophrenia and dopamine modulation in the prefrontal cortex. *Nat Rev Neurosci.* 2008;9:696–709.
49. Friston K, Kiebel S. Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci.* 2009;364:1211–1221.
50. Yedidia JS, Freeman WT, Weiss Y. Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Trans Inf Theory.* 2005;51:2282–2312.
51. Smolensky P. *Information Processing in Dynamical Systems: Foundations of Harmony Theory.* Boulder, CO: University of Colorado Boulder Computer Science Department; 1986.
52. Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA.* 1982; 79:2554–2558.
53. Dayan P, Hinton GE, Neal RM, Zemel RS. The Helmholtz machine. *Neural Comput.* 1995;7:889–904.
54. Del Arco A, Mora F. NMDA and AMPA/kainate glutamatergic agonists increase the extracellular concentrations of GABA in the prefrontal cortex of the freely moving rat: modulation by endogenous dopamine. *Brain Res Bull.* 2002;57:623–630.
55. Loh M, Rolls ET, Deco G. A dynamical systems hypothesis of schizophrenia. *PLoS Comput Biol.* 2007;3:e228.
56. Anticevic A, Lisman J. How can global alteration of excitation/inhibition balance lead to the local dysfunctions that underlie schizophrenia? *Biol Psychiatry.* 2017;81:818–820.
57. Rogasch NC, Daskalakis ZJ, Fitzgerald PB. Cortical inhibition, excitation, and connectivity in schizophrenia: a review of insights from transcranial magnetic stimulation. *Schizophr Bull.* 2014;40:685–696.
58. Grace AA. Dysregulation of the dopamine system in the pathophysiology of schizophrenia and depression. *Nat Rev Neurosci.* 2016;17:524–532.
59. Hoffman RE, McGlashan TH. Using a speech perception neural network computer simulation to contrast neuroanatomic versus neuromodulatory models of auditory hallucinations. *Pharmacopsychiatry.* 2006;39(Suppl 1):S54–S64.
60. Ffytche DH. The hodology of hallucinations. *Cortex.* 2008;44:1067–1083.
61. Upthegrove R, Broome MR, Caldwell K, Ives J, Oyeboode F, Wood SJ. Understanding auditory verbal hallucinations: a systematic review of current evidence. *Acta Psychiatr Scand.* 2016;133:352–367.
62. Thaker GK, Ross DE, Buchanan RW, Adami HM, Medoff DR. Smooth pursuit eye movements to extra-retinal motion signals: deficits in patients with schizophrenia. *Psychiatry Res.* 1999;88:209–219.
63. Frith CD. *The Cognitive Neuropsychology of Schizophrenia.* Hillsdale, NJ, England: Lawrence Erlbaum Associates; 1992.
64. Blakemore SJ, Smith J, Steel R, Johnstone CE, Frith CD. The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychol Med.* 2000;30:1131–1139.
65. Shergill SS, Samson G, Bays PM, Frith CD, Wolpert DM. Evidence for sensory prediction deficits in schizophrenia. *Am J Psychiatry.* 2005;162:2384–2386.
66. Waters F, Allen P, Aleman A, et al. Auditory hallucinations in schizophrenia and nonschizophrenia populations: a review and integrated model of cognitive mechanisms. *Schizophr Bull.* 2012;38:683–693.
67. Fine C, Gardner M, Craigie J, Gold I. Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cogn Neuropsychiatry.* 2007;12:46–77.
68. Klosterkötter J. The meaning of basic symptoms for the genesis of the schizophrenic nuclear syndrome. *Jpn J Psychiatry Neurol.* 1992;46:609–630.
69. Ratcliffe M. *Real Hallucinations: Psychiatric Illness, Intentionality, and the Interpersonal World.* Cambridge, MA: MIT Press; 2017.
70. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. *Front Psychiatry.* 2013;4:47.
71. Friston KJ. Hallucinations and perceptual inference. *Behav Brain Sci.* 2005;28:764–766.
72. Jardri R, Denève S. Computational models of hallucinations. In: Jardri R, Cachia A, Thomas P, Pins D, eds. *The Neuroscience of Hallucinations.* New York, NY: Springer; 2013:289–313.
73. Javitt DC, Sweet RA. Auditory dysfunction in schizophrenia: integrating clinical and basic features. *Nat Rev Neurosci.* 2015;16:535–550.
74. Teufel C, Subramaniam N, Dobler V, et al. Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc Natl Acad Sci U S A.* 2015;112:13401–13406.
75. Schmack K, Gómez-Carrillo de Castro A, Rothkirch M, et al. Delusions and the role of beliefs in perceptual inference. *J Neurosci.* 2013;33:13701–13712.
76. Jardri R, Denève S. Circular inferences in schizophrenia. *Brain.* 2013;136:3227–3241.
77. Denève S, Jardri R. Circular inference: mistaken belief, misplaced trust. *Curr Opin Behav Sci.* 2016;11:40–48.
78. Jardri R, Duverne S, Litvinova AS, Denève S. Experimental evidence for circular inference in schizophrenia. *Nat Commun.* 2017;8:14218.
79. Teufel C, Kingdon A, Ingram JN, Wolpert DM, Fletcher PC. Deficits in sensory prediction are related to delusional ideation in healthy individuals. *Neuropsychologia.* 2010;48:4169–4172.
80. Brown H, Adams RA, Parees I, Edwards M, Friston K. Active inference, sensory attenuation and illusions. *Cogn Process.* 2013;14:411–427.
81. Benrimoh D, Parr T, Vincent P, Adams RA, Friston K. Active inference and auditory hallucinations. *Computational Psychiatry.* 2018;1–22.
82. McCarthy-Jones S, Trauer T, Mackinnon A, Sims E, Thomas N, Copolov DL. A new phenomenological survey of auditory hallucinations: evidence for subtypes and implications for theory and practice. *Schizophr Bull.* 2014;40:231–235.
83. Jardri R, Thomas P, Delmaire C, Delion P, Pins D. The neurodynamic organization of modality-dependent hallucinations. *Cereb Cortex.* 2013;23:1108–1117.
84. Johns LC, Kompus K, Connell M, et al. Auditory verbal hallucinations in persons with and without a need for care. *Schizophr Bull.* 2014;40(Suppl 4):S255–S264.
85. Hill K, Varese F, Jackson M, Linden DE. The relationship between metacognitive beliefs, auditory hallucinations, and hallucination-related distress in clinical and non-clinical voice-hearers. *Br J Clin Psychol.* 2012;51:434–447.
86. Horga G, Schatz KC, Abi-Dargham A, Peterson BS. Deficits in predictive coding underlie hallucinations in schizophrenia. *J Neurosci.* 2014;34:8072–8082.
87. Thuné H, Recasens M, Uhlhaas PJ. The 40-Hz auditory steady-state response in patients with schizophrenia: a meta-analysis. *JAMA Psychiatry.* 2016;73:1145–1153.

88. Powers AR, Mathys C, Corlett PR. Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science*. 2017;357:596–600.
89. Cassidy CM, Balsam PD, Weinstein JJ, et al. A perceptual inference mechanism for hallucinations linked to striatal dopamine. *Curr Biol*. 2018;28:503–514.e4.
90. Kapur S. Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. *Am J Psychiatry*. 2003;160:13–23.
91. Martin JR, Pacherie E. Out of nowhere: thought insertion, ownership and context-integration. *Conscious Cogn*. 2013;22:111–122.
92. Turner MS, Simons JS, Gilbert SJ, Frith CD, Burgess PW. Distinct roles for lateral and medial rostral prefrontal cortex in source monitoring of perceived and imagined events. *Neuropsychologia*. 2008;46:1442–1453.
93. Mondino M, Haesebaert F, Poulet E, Suaud-Chagny MF, Brunelin J. Fronto-temporal transcranial Direct Current Stimulation (tDCS) reduces source-monitoring deficits and auditory hallucinations in patients with schizophrenia. *Schizophr Res*. 2015;161:515–516.
94. Synofzik M, Vosgerau G, Newen A. I move, therefore I am: a new theoretical framework to investigate agency and ownership. *Conscious Cogn*. 2008;17:411–424.
95. Sato A, Yasuda A. Illusion of sense of self-agency: discrepancy between the predicted and actual sensory consequences of actions modulates the sense of self-agency, but not the sense of self-ownership. *Cognition*. 2005;94:241–255.
96. Vosgerau G, Newen A. Thoughts, motor actions, and the self. *Mind & Language*. 2007;22:22–43.
97. Mullins S, Spence SA. Re-examining thought insertion. Semi-structured literature review and conceptual analysis. *Br J Psychiatry*. 2003;182:293–298.
98. Campbell J. Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist*. 1999;82:609–625.
99. Gerrans P. *The Measure of Madness: Philosophy of Mind, Cognitive Neuroscience, and Delusional Thought*. Cambridge, MA: MIT Press; 2014.
100. Walsh E, Oakley DA, Halligan PW, Mehta MA, Deeley Q. Brain mechanisms for loss of awareness of thought and movement. *Soc Cogn Affect Neurosci*. 2017;12:793–801.
101. Jardri R, Delevoye-Turrell Y, Lucas B, et al. Clinical practice of rTMS reveals a functional dissociation between agency and hallucinations in schizophrenia. *Neuropsychologia*. 2009;47:132–138.
102. Stephens GL, Graham G. *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge, MA: MIT Press; 2000.
103. Frith C, Rees G, Friston K. Psychosis and the experience of self: brain systems underlying self-monitoring. *Ann N Y Acad Sci*. 1998;843:170–178.
104. Alderson-Day B, Lima CF, Evans S, et al. Distinct processing of ambiguous speech in people with non-clinical auditory verbal hallucinations. *Brain* 2017;140:2475–89.
105. Langland-Hassan P. Fractured phenomenologies: thought insertion, inner speech, and the puzzle of extraneity. *Mind & Language*. 2008;23:369–401.
106. Sarbin TR. The concept of hallucination. *J Pers*. 1967;35:359–380.