RUNNING HEAD: MISATTRIBUTION MODELS

Misattribution Models (II): Source Monitoring in Hallucinating Schizophrenia Subjects

(Chapter 10. In Section 2: Cognitive Models of Hallucinations)

in

The Neuroscience of Hallucinations.

Renaud Jardri, Delphine Pins, Arnaud Cachia & Pierre Thomas (Eds.) Springer

by

Todd S. Woodward[1,2,*] & Mahesh Menon[3,4]

[1] Department of Psychiatry, University of British Columbia, Vancouver, Canada

[2] Department of Research, BC Mental Health & Addictions Research Institute, Vancouver Canada

[3] Schizophrenia Program and PET Centre, Centre for Addiction & Mental Health, Toronto, Canada

[4] Department of Psychiatry, University of Toronto, Toronto, Canada

[*] Corresponding author. Please address correspondence to Todd S. Woodward, Ph.D., BC Mental Health & Addictions Research Institute, Room 116, 3rd Floor, 938 W. 28th Avenue, Vancouver, B.C., Canada. V5Z 4H4. e-mail: toddswoodward@gmail.com. Phone: 604-875-2000 x 4724. Fax: 604-875-3871.

Abstract

Misattribution models of auditory vocal hallucinations hold that the cognitive operations involved in generating internal thoughts are biased in some fashion, such that these self-generated thoughts are experienced as nonself-generated voices. Source monitoring refers to a set of normal cognitive operations that are invoked to determine the context and conditions under which a memory episode was encoded. Source monitoring paradigms are relevant to misattribution models of hallucinations when two of the monitored sources are inner (self) and outer (e.g., other) because they are assumed to share overlapping cognitive operations that lead to misattributing self-generated cognitive events to an external source (i.e., externalizations). In this chapter we review how certain types of inner/outer source monitoring errors can inform misattribution models of hallucinations, and bring forward methodological issues related to task design, data analysis, and group comparisons. We review the theoretical accounts of the cognitive operations shared between hallucinations and source monitoring, and the functional neuroimaging data that bears on those models. We conclude that sufficient evidence for an association between hallucinations and externalization errors in source monitoring has been demonstrated, and that the neuroimaging results support hyperactivation of voice selective cortical regions (e.g., the superior temporal gyrus) as underlying the cognitive operations shared between externalization errors in source monitoring and hallucinations. We provide suggestions for more comprehensive testing of misattribution models using more refined cognitive paradigms.

KEYWORDS: HALLUCINATIONS; SCHIZOPHRENIA; SOURCE MONITORING; MULTINOMIAL MODELING; MISATTRIBUTION MODELS

ABBREVIATIONS: ISG (Inner self-generated); ING (Inner non-self-generated); OSG (Outer self-generated); ONG (Outer non-self-generated); MCT (Meta-Cognitive Training)

Introduction

Hallucinations refer to sensory experiences occurring in the absence of the corresponding external stimulus. Although they can occur in any of the sensory modalities, auditory hallucinations, typically voices, are the most common form in schizophrenia. Phenomenologically, auditory hallucinations may be perceived as originating from sources external to the person, or as being heard 'inside the head', with the latter category traditionally categorized as 'pseudohallucinations' (Jaspers, 1968/1912). Most research on auditory hallucinations has not attempted to differentiate between hallucinations depending on their source, although some researchers have speculated that there might be different neural or cognitive mechanisms involved in both (Hunter, 2004; Plaze et al., 2011), while others have suggested that patients may not be able to reliably localize the source of their voices (Nayani & David, 1996). Importantly, both types of auditory hallucinations involve the experience of voices that are perceived as originating from some external source (even if they are 'heard' inside the head), and not seen by the patient as being the result of their own thought process.

Hallucinations have long been theorized to be a form of verbal thought. Early psychological investigations found that thinking is usually accompanied by subvocalizations and corresponding electrical activity in the vocal musculature (Pinter, 1913), and subsequently it was discovered that auditory hallucinations were accompanied by similar subvocalizations (Gould, 1948). Furthermore, it was noted that tasks that disrupted subvocalization reduced the hallucination (Green & Kinsbourne, 1989). Thus, when attempting to understand the cognitive underpinnings of hallucinations, consideration must be given to how verbal thoughts could transform into

hallucinations. Source monitoring and misattribution models have been employed as a methodology for providing insight into the cognitive operations that could lead to this transformation.

What is a Misattribution Model?

Cognitive models of hallucinations have suggested that hallucinations might occur as a result of biases in 'normal' cognitive processes involved in speech perception or source monitoring such as auditory imagery (Mintz & Alpert, 1972), self-monitoring deficits (Frith & Done, 1988) and source monitoring deficits (Keefe, Arnold, Bayen, & Harvey, 1999). Misattribution models can be considered one type of cognitive model of hallucinations. To introduce misattribution models we refer to Larøi and Woodward (Larøi & Woodward, 2007), and particularly Table 1 from that paper (presented here in a modified form; Table 1), which focuses on two important phenomenological dimensions in hallucinations: (1) the self-generated/non-self-generated dimension, and (2) the inner/outer dimension (Stephane, Thuras, Nasrallah, & Georgopoulos, 2003). The former dimension refers to the perceived/subjective generating origin of a given cognitive event, and the latter to the localization of the cognitive event in space. A combination of these two dimensions gives rise to four different types of cognitive events. Inner, self-generated (ISG) cognitive events are, of course, healthy and normal. All hallucinations originate as inner, self-generated cognitive events, but are somehow altered such that they are experienced as one of the other three cells in Table 1 (OSG, ONG, or ING). That is, either because of changes in the person's explanation of the event's perceived subjective origin (i.e., becomes less "self generated") or due to changes in the event's subjective spatial location (i.e., becomes less internal or inner), or a combination of both, the inner, self-generated event is transformed into a hallucination. In this context, hallucinations may be viewed as inner, self-generated events that are misattributed, hence the relevance to misattribution models of hallucinations.

This model assumes that two basic steps underlie the onset of hallucinations: (1) the *alienation* of inner, self-generated events and (2) the *misattribution* of inner, self-generated events to some specific origin/location combination. The first type of cognitive process, alienation, involves loss of the cognitive representations that code the inner, self-generated nature of what will become a hallucination. In other words, an inner, self-generated event is (subjectively speaking) not clearly experienced as inner, self-generated, or both. The alienation process is covered in detail in other sources (Larøi & Woodward, 2007; Waters, Woodward, Allen, Aleman, & Sommer, 2010), and although the alienation step is a necessary component of misattribution models, it is not the essential step, with the essential step being the misattribution step. A misattribution model, from a cognitive perspective, is a proposed set of cognitive operations that have been affected in hallucinating patients, and have resulted in one of the shifts from the ISG cell to the OSG, ONG, or ING cell, as is depicted in Table 1.

What is Source Monitoring?

Source monitoring is a very general term that refers to the ability to distinguish the wide range of variables that specify the context and conditions under which a memory episode was encoded (Johnson, Hashtroudi, & Lindsay, 1993). Source information is considered retrievable in memory on the basis of the qualitative and quantitative characteristics of the memories themselves. A number of different studies provide evidence showing that people do indeed use associated perceptual, semantic and other information to remember the source of memories (reviewed in Johnson et al., 1993).

Central to Johnson and co-workers approach is the idea that people typically do not retrieve an abstract tag or label that specifies a memory's source, but rather that memory records are evaluated and attributed to particular sources through decision processes activated during

remembering, and source monitoring itself refers to the set of processes involved in making the attributions about the source of memories, knowledge and beliefs. Among the most important memory characteristics are records of perceptual information (e.g. sound and colour), contextual data (e.g., spatial and temporal information), semantic detail, affective information and cognitive operations (e.g. records of organising, elaborating, retrieving, and identifying) that were established when the memory was formed. Source monitoring decisions capitalize on average differences in characteristics of memories from various sources. For example, compared to memories for imagined events, memories for perceived events tend to include more perceptual, spatial, temporal semantic and affective information, and less information about cognitive operations. Consequently, a memory with, say, a great deal of visual and spatial detail and very little detail regarding generating cognitive operations is more likely to be remembered as being externally derived, regardless of its origin. These processes may be relatively automatic 'heuristic processes' or carried out via slower, controlled, 'systematic processes', which may occur consciously. Similarly, a memory with particular contextual information (e.g. "I remember talking about that while we were in the restaurant") might help to identify the 'source' of that memory (in this case, who the speaker was).

The source monitoring framework is an extension of the reality monitoring framework (Johnson & Raye, 1981). Reality monitoring refers to the ability to discriminate memories of internally generated information from memories of externally derived information, such as the ability to differentiate between thoughts and imaginations from memories of perceived events. In the same way as was mentioned for source monitoring above, in reality monitoring, the source of the memories should be able to be distinguished on the basis of their contents. Although the type of source monitoring tasks used to test misattribution models of hallucinations would strictly fall

under the rubric of reality monitoring, the term source monitoring has been used in almost all studies.

Source Monitoring Errors that can be Applied to Misattribution Models of Hallucinations

Source monitoring is one of the leading cognitive paradigms for the study of auditory hallucinations, because auditory hallucinations can be conceptualized as the assignment of internally generated mental episodes to an external source; therefore, an association between hallucinations and a tendency to misremember an internally generated event as originating from an external source would suggest overlapping cognitive operations for the two. That is to say, the bearing of source monitoring evidence on misattribution theories of hallucinations need not be based on an assumption that hallucinations require a memory component, but rather that overlapping cognitive operations underlie both self-other source monitoring and the presence of hallucinations. Evidence from source monitoring tasks that bears most directly on misattribution models of hallucinations involves the self-other (or inner/outer) source distinction bring made in memory. Research using self-other source stimuli that are not memory dependent (Allen et al., 2004; Blakemore, Smith, Steel, Johnstone, & Frith, 2000; Johns et al., 2001; Stephane, Kuskowski, McClannahan, Surerus, & Nelson, 2009) also provides evidence that bears on misattribution models. Some evidence from that body of research suggest that self-other difficulties are seen in hallucinating patients, suggesting overlapping cognitive operations with hallucinations, but these studies are not reviewed here because they do not fit into the memory-based source monitoring framework put forward by Johnson and co-workers (this being the topic of the current chapter).

In source monitoring studies on schizophrenia that group patients on presence/absence of hallucinations, a difference between the groups on a tendency to misremember an internally generated event as originating from an external source is consistently reported (Bentall, Baker, &

Havers, 1991; Brébion et al., 2000; Brunelin et al., 2006; Costafreda, Brebion, Allen, McGuire, & Fu, 2008; Franck et al., 2000; Woodward, Menon, & Whitman, 2007), including for healthy subjects who report hearing voices (Larøi, Van Der Linden, & Marczewski, 2004; Rankin & O'Carroll, 1995), and even siblings of hallucinating patients (Brunelin et al., 2007) although failures to replicate have also been reported (e.g., Seal, Crowe, & Cheung, 1997). Table 2 lists details regarding sample size and experimental design for source monitoring studies where hallucinating patients were compared to non-hallucinating patients on externalization-type memory errors. Evidence for this externalization bias (ISG→ONG variety from Table 1) has been confirmed in two reviews of the literature (Ditman & Kuperberg, 2005; Seal, Aleman, & McGuire, 2004). Although apparently reliable, observation of the association between the externalization bias and hallucinations can be affected by a number of methodological issues, including the nature of the source monitoring task, data analysis method, and group comparisons carried out. These issues will be addressed in more detail in the next section.

Theoretical Accounts

Although the association of externalization biases (misattributions of the ISG→ONG variety) and hallucinations suggests overlapping cognitive operations between inner/outer source confusion in memory and hallucinations, this evidence alone does not clearly identify those cognitive operations. However, a number of existing theoretical accounts are directly relevant to the identification of these overlapping cognitive operations. Bentall (1990) first proposed the importance of top-down processes in the overlap between hallucinations and source monitoring by putting forward a "metacognitive skills" deficit. This was taken one step further by Frith's theoretical account of hallucinations (Frith, 1992; Frith et al., 1995), which assumed that receptive centres of the brain (involved with speech perception) receive afferent feedback from speech

generation centres of the brain, and that this feedback is the internal signal indicating that a thought that is verbalized has been internally generated. The theory suggested that schizophrenia may be characterized by a disconnection, whereby information about willed intentions does not reach self-monitoring centers, causing self-generated material to be not recognized as self-generated (Shergill, Bullmore, Simmons, Murray, & McGuire, 2000). Although Frith stated (Frith, 1992, p. 84) that this account would only clearly explain certain types of hallucinations, particularly "hearing one's own thoughts spoken aloud" (traditionally referred to as "gedankenlautwerden"), it has gained much traction, and follow-up studies proposed a biological disconnection, involving loss of moderating frontal cortex influence on the superior temporal cortex. Neuroimaging studies, particularly fMRI studies, have provided evidence for this functional disconnection (Ford, Mathalon, Whitfield, Faustman, & Roth, 2002; McGuire et al., 1995; McGuire et al., 1996), showing reduced frontal activity as well as reduced functional connectivity between the dorsolateral prefrontal cortex and the superior temporal cortex. This was proposed to result in hyperactivity of voice-sensitive regions.

Although Frith's theory is typically considered a top-down account, due to the involvement of self-monitoring, it includes both top-down and bottom-up components. Both the frontal cortex (top-down) and temporal cortex (bottom-up) aspects of the disconnection extension of this theory (McGuire et al., 1995; McGuire et al., 1996) could explain the observed association between externalization errors in source monitoring and hallucinations. For example, a degraded top-down "generating-thoughts" signal would bias any self-generated cognitive event to be interpreted as other-generated. A different possible cause based on bottom-up processes, that would produce the same externalization bias in source memory, is the addition of vivid verbal sensations to private thoughts due to hyperactivity of auditory perception regions of the brain, causing them to be 'perceived' as external.

Reviews of the literature have found substantial evidence in support of the bottom-up account, concluding that hyperactivity of the superior temporal gurus (STG) is an essential component of the "over-perceptualization" account of hallucinations (Allen, Larøi, McGuire, & Aleman, 2008). Hyperactivity of voice-selective regions of the STG, even for silent thought, has been implicated in a number of empirical and theoretical works on hallucinations (Ford, Roach, Faustman, & Mathalon, 2007; Northoff & Qin, 2011; Rapin et al., in revision), and it has been demonstrated that repeated transcranial magnetic stimulation (rTMS) applied to voice-selective cortical regions reduces the intensity of hallucinations (Hoffman et al., 2003; Jardri et al., 2009; Vercammen, Knegtering, Liemburg, den Boer, & Aleman, 2010). Correspondingly, and not surprisingly, these voice-selective cortical regions activate when patients report hearing hallucinations (Sommer et al., 2008; Suzuki, Yuasa, Minabe, Murata, & Kurachi, 1993) and are hyperactive during resting state for those patients (Dierks et al., 1999). Also supporting this account is the observation that voice hearers report increased vividness and loudness for normal thoughts (Moritz & Laroi, 2008).

Relative to the strong evidence for the over-perceptualization account of hallucinations, direct evidence for degraded top-down "generating-thoughts" signal is lacking, and top-down dysfunction is considered secondary to bottom-up overactivity (Allen et al., 2008). Relating this back to Frith's theory and it's biological extension to a disconnection syndrome, it been pointed out that disconnection is not required to produce hyperactivity in voice-selective cortical regions, and that a "breakaway" speech perception network could occur spontaneously, possibly manifesting as disconnection with frontal regions (David, 1994; Hoffman, 2010; Hoffman, 1999). Thus, hyperactivity in the voice-selective temporal cortex regions could provide fertile ground for hallucinations, and would also lead to externalization errors in source monitoring experiments, even in the absence of a top-down deficit or a disconnection with frontal regions.

Some theoretical accounts have proposed biochemical abnormalities that may underlie the bottom-up "over-perceptualization" account of hallucinations, and the associated externalization errors in source monitoring experiments. For example, a hyperdopaminergic state (Kapur, 2003) experienced during episodes of psychosis may result in these internal percepts appearing more "salient". This could lead to an increase in the perceptual qualities of thoughts mediated by hyperactivity in voice-selective cortical regions, biasing thought to be perceived as externally generated, and leading to externalization errors in source monitoring experiments. Other biological substrates that have been proposed to lead to hypersalience of percepts, hyperactivity of voice-selective cortical regions, or hallucinations are: N-methyl-D-aspartate receptor (NMDAR) dysfunction (Stephan, Friston, & Frith, 2009; Stone et al., 2011) and GABA (γ-aminobutyric acid)-ergic receptor dysfunction (Rotarska-Jagiela et al., 2010).

Although hyperactivity of voice-selective cortical regions would explain the overlap between externalizations in source monitoring and hallucinations (this being the topic of the current chapter) this is presumably not a full explanation for the manifestation of hallucinations. Metacognitive or belief-based influences are also likely play a role (Allen et al., 2008; Bentall, 1990; Moritz & Laroi, 2008). Most people acknowledge the presence of occasional intrusive thoughts with negative content, but in certain psychiatric conditions (e.g., OCD and hallucinations), individuals may be more distressed by the presence of these thoughts. In patients with hallucinations, this may result in feelings of reduced controllability of their thoughts as well as reduced internality (i.e., the feeling that the thought *must* have been from an outside source due to its negative content). These maladaptive metacognitive beliefs could in turn lead to strategies such as thought suppression, possibly having the contradictory rebound effect, resulting in an increased frequency of such negative thoughts, and reduced feelings of controllability or internality of the thoughts. This may result in thoughts with negative content considered to be originating from an

external source - a set of beliefs which might exacerbate the source monitoring deficits outlined earlier (Baker & Morrison, 1998; Moritz & Laroi, 2008; Morrison, Haddock, & Tarrier, 1995) - and possibly be exacerbated by dysfunction in the brain centered underlying emotion (Allen et al., 2008).

Methodological Considerations

Which Source Monitoring Task to Use?

Typically, items in a source monitoring task are provided by an external source, and others are generated internally. Most source-monitoring studies ask participants to verbally generate a word when attempting to produce an instantiation of the internal source. One interpretational difficulty associated with this methodology is that although the event is self generated, it contains both inner *and* outer localization qualities. Specifically, the generation of the word is, indeed, an inner event, but the production of the word also leads to stimulation of sensory organs, thereby adding outer localization qualities. Thus, a purely inner, self-generated event that seems a basic requirement for the study of alienation is rarely used. Instead, the self-generated event is typically characterized by a mixture of inner and outer qualities.

However, even accepting the assumption that the experimental conditions do reflect pure inner and outer sources, most studies include only *one* internal and *one* external source (Baker & Morrison, 1998; Bentall et al., 1991; Bentall & Slade, 1985; Ensum & Morrison, 2003; Johns & McGuire, 1999; Larøi et al., 2004; Morrison & Haddock, 1997; Rankin & O'Carroll, 1995; Seal et al., 1997). This restricted set of conditions limits the specificity of conclusions in the following ways: (1) only shifts from the ISG to ONG cells of Table 1 can be tested, leaving auditory hallucinations involving misattributions to other cells (e.g, ISG→ING and ISG→OSG) unstudied. (2) Other types of misattributions cannot be used as control conditions for the ISG→ONG type, for

example, ISG→ING, ISG→OSG, or confusion of two external sources (the latter was used in some studies as a control; Brébion et al., 2000; Keefe, Arnold, Bayen, McEvoy, & Wilson, 2002; Woodward et al., 2007). (3) Assuming alienation has taken place for a recognized item, cognitive operations associated with ISG→ING and ISG→OSG misattributions would be merged with those underlying ISG→ONG misattribution errors, because the ONG source is the only misattribution response possible following alienation, potentially leading to interpretational difficulties.

Another important consideration for source monitoring studies, due to their focus on error types, is the number of items per condition/source combination. As can be seen in Table 2, the number of items per condition/source combination exceeded 16 in two studies only. The estimated sample error associated with a mean increases as the number of items per condition decreases (a consequence of the central limit theorem), therefore concerns over the number of items per condition/source combination are especially pronounced when specific error types (e.g., externalizations and internalizations) are the measures of interest.

In summary, in order to integrate the full range of origin/source combinations into the experimental design of source monitoring investigations of auditory hallucinations, source-monitoring studies would ideally include (1) include enough trials to measure reliably the error types of theoretical interest, (2) include pure internal generation events (i.e., thoughts, images) that are not confounded by the external location quality that results from saying a word, (3) use experimental designs that allow investigation of biases of the ISG→ING and ISG→OSG type, and (4) integrate built-in experimental controls such as confusion of two external sources.

Which Data Analysis Method to Use?

An interpretational challenge associated with source monitoring data is to distinguish between, on one hand, externalization errors on which an internal event has been transformed to an

external event by the same cognitive processes that produce hallucinations, and, on the other hand, externalization errors on which the subjects simply guessed that the source was external. Such guessing events may be common. For example when subjects notice that they are recognizing too few items from the (less memorable) external source, they tend to compensate by increasing the number of external-source guesses (Batchelder & Riefer, 1990). This occurs on any type of trial, whether or not the presented item was new, internally generated, or externally generated. When these guessing processes take place in response to internally generated but forgotten items, these errors will be summed along with true externalization biases in the count of externalizations. If such confounds are not properly accounted for, inaccurate conclusions can be reached; for example, a true externalization bias in the hallucinating-group may be masked by increases in external-source guesses in the comparison-group. Using appropriate statistical techniques, distinct cognitive processes such as pure guessing and cognitive biases can be disentangled. Therefore, in order to accurately quantify true externalizations that are of interest to the study of hallucinations, increases in these strategic "external" guesses should be partialled out.

The application of multinomial modeling to the study of the positive symptoms of schizophrenia has been carried out under the motivation of disentangling guessing from true biases. In fact, the overall goal of multinomial modeling is to eliminate the contribution of guesses from the parameter estimates. However, primarily due to the restrictions in the allowable number of parameters in a model, source-specific biases or guessing estimates typically cannot be estimated (see Menon & Woodward, 2007 for a detailed discussion of issues related to multinomial modeling analysis and other analysis methods; Woodward & Menon, 2010).

An alternative is to consider the common practice in memory experiments and other standard neuropsychological measures of memory of computing a "corrected recognition" index, which involves subtracting the false positive rate from the recognition rate in order to provide a

measure of recognition rate not inflated by guesses (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Nelson et al., 2003). A similar procedure that will allow guesses to be subtracted out of measurement of error rates involves using analysis of covariance (ANCOVA), where the restrictions that apply to multinomial modelling in terms of the number of estimable parameters do not apply (see Woodward & Menon, 2010). Guessing rates can be estimated by false positive rates, which are readily available for all sources, and are simply the number of trials on which a particular source response is given in response to new items. For example, for externalization errors (i.e., the number of times a person responded with an external source for an internally generated item) we subtract out guesses by using the number of trials on which that subject responded with an external source in response to *new* items as a covariate in ANCOVA. The same strategy could be used for most error types of interest, and comparison between groups is also carried out naturally in the ANCOVA framework. This methodology also requires a significance test of each covariate, checking the sphericity assumption, and checking the homogeneity of regression assumption, all of which are readily available with standard ANCOVA procedures (Woodward & Menon, 2010).

Which Groups to Compare?

An understanding of the cognitive processes involved in the predisposition to hallucinations and those associated with the immediate experience of hallucinations ideally requires samples of current hallucinators (e.g., with a daily frequency of occurrence), past hallucinators (those who showed the symptom in the past but no longer show it), schizophrenia patients who have never had hallucinations, as well as a psychiatric control group (of non-schizophrenia patients without hallucinations) and non-psychiatric controls. In order to examine the specificity of the symptom to schizophrenia, other studies may also wish to compare hallucinations in schizophrenia with those seen in other syndromes (e.g. depression). Although it is usually not practical to collect a

comprehensive dataset that includes all of these groups, at minimum, groups of hallucinating schizophrenia patients must be compared to nonhallucinating schizophrenia patients (not only to healthy controls) to provide evidence that the proposed cognitive underpinnings of hallucinations change as the symptoms change, and are state aspects of the illness instead of, or in addition to, trait aspects. In other words, evidence bearing on investigation of cognitive processes underlying hallucinations is weak when derived from studies comparing a group of patients diagnosed with schizophrenia to a control group without consideration for variation in the manifestation of hallucinations within that patient group, particularly when using a healthy control group that differs from the patients group on a number of variables that are confounded with diagnosis, such as general cognitive abilities, medication, and education, to name a few.

Summary

The basis of this review is that inner/outer source monitoring memory paradigms are relevant to misattribution models of hallucinations because they are assumed to share overlapping cognitive operations that lead to misattributing self-generated cognitive events to an external source (i.e., externalizations). Our review of the source monitoring studies on schizophrenia for which patients were grouped on presence/absence of hallucinations demonstrated an increased frequency of misattributing internally generated events to an external source for the hallucinating patients. Our review of the theoretical and empirical accounts of these shared cognitive operations pointed to bottom-up hyperactivation of voice selective cortical regions (e.g., the superior temporal gyrus), and the resultant addition of vivid voice-based perceptual information to internally generated cognitive events, as the most parsimonious explanation for the association between externalization errors in source monitoring and hallucinations. We also outline how this could contribute to hallucinations alongside top-down, belief based cognitive operations. Finally, we provide

suggestions for more comprehensive testing of misattribution models using more refined cognitive paradigms, and methodological considerations for task design and data analytic strategies.

Cognitive models of hallucinations such as misattribution models have played an important role in developing novel treatments of auditory hallucinations, including cognitive behaviour therapy for voices which are focussed on helping people understand the emotional implications and effects of the associated beliefs (see Van der Gaag, 2006 for a comprehensive model), as well as metacognitive training (MCT; Moritz & Woodward, 2007a; Moritz & Woodward, 2007b), aimed at making people aware of cognitive biases and their relationship to symptomatology. Although the methodological sections of this chapter have pointed out limitations to this body of work, the empirical evidence for an association between externalization errors in source monitoring experiments and hallucinations is currently one of the most well-replicating findings in cognitive neuropsychiatric investigations into the symptoms of schizophrenia. We hope that future work in this area can benefit from the methodological and theoretical suggestions put forward in this article.

References

Allen, P., Larøi, F., McGuire, P. K., & Aleman, A. (2008). The hallucinating brain: A review of structural and functional neuroimaging studies of hallucinations. *Neurosci Biobehav Rev, 32*, 175-191.

Allen, P. P., Johns, L. C., Fu, C. H. Y., Broome, M. R., Vythelingum, G. N., & McGuire, P. K. (2004). Misattribution of external speech in patients with hallucinations and delusions. *Schizophrenia Research, 69*, 277-287.

Baker, C. A., & Morrison, A. P. (1998). Cognitive processes in auditory hallucinations: Attributional biases and metacognition. *Psychological Medicine, 28*, 1199-1208.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97*, 548-564.

Bentall, R. P. (1990). The illusion of reality: A review and integration of psychological research on hallucinations. *Psychological Bulletin, 107*, 82-95.

Bentall, R. P., Baker, G. A., & Havers, S. (1991). Reality monitoring and psychotic hallucinations. *British Journal of Clinical Psychology, 30*, 213-222.

Bentall, R. P., & Slade, P. D. (1985). Reality testing and auditory hallucinations: A signal detection analysis. *British Journal of Clinical Psychology, 24*, 159-169.

Blakemore, S. J., Smith, J., Steel, R., Johnstone, C. E., & Frith, C. D. (2000). The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: Evidence for a breakdown in self-monitoring. *Psychol Med, 30*, 1131-1139.

Brébion, G., Amador, X., David, A., Malaspina, D., Sharif, Z., & Gorman, J. M. (2000). Positive symptomatology and source-monitoring failure in schizophrenia-an analysis of symptom-specific effects. *Psychiatry Research, 95*, 119-131.

Brunelin, J., Combris, M., Poulet, E., Kallel, L., D'Amato, T., Dalery, J., & Saoud, M. (2006). Source monitoring deficits in hallucinating compared to non-hallucinating patients with schizophrenia. *Eur Psychiatry, 21*, 259-261.

Brunelin, J., d'Amato, T., Brun, P., Bediou, B., Kallel, L., Senn, M., Poulet, E., & Saoud, M. (2007). Impaired verbal source monitoring in schizophrenia: An intermediate trait vulnerability marker? *Schizophrenia Research, 89*, 287-292.

Costafreda, S. G., Brebion, G., Allen, P., McGuire, P. K., & Fu, C. H. (2008). Affective modulation of external misattribution bias in source monitoring in schizophrenia. *Psychol Med, 38*, 821-824.

David, A. S. (1994). The neuropsychological origin of auditory hallucinations. In A. S. David & J. C. Cutting (Eds.), *The neuropsychology of schizophrenia* (pp. 269-313). Hillside, NJ: Erlbaum.

Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., & Singer, W. (1999). Activation of heschl's gyrus during auditory hallucinations. *Neuron, 22*, 615-621.

Ditman, T., & Kuperberg, G. R. (2005). A source-monitoring account of auditory verbal hallucinations in patients with schizophrenia. *Harvard Review of Psychiatry, 13*, 280-299.

Ensum, I., & Morrison, A. P. (2003). The effects of focus of attention on attributional bias in patients experiencing auditory hallucinations. *Behaviour Research and Therapy, 41*, 895-907.

Ford, J. M., Mathalon, D. H., Whitfield, S., Faustman, W. O., & Roth, W. T. (2002). Reduced communication between frontal and temporal lobes during talking in schizophrenia. *Biological Psychiatry, 51*, 485-492.

Ford, J. M., Roach, B. J., Faustman, W. O., & Mathalon, D. H. (2007). Synch before you speak: Auditory hallucinations in schizophrenia. *American Journal of Psychiatry, 164*, 458-466.

Franck, N., Rouby, P., Daprati, E., Dalery, J., Marie-Cardine, M., & Georgieff, N. (2000). Confusion between silent and overt reading in schizophorenia. *Schizophrenia Research, 41*, 357-364.

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia*. Hillsdale, NJ: Lawrence Erlbaum.

Frith, C. D., & Done, D. J. (1988). Towards a neuropsychology of schizophrenia. *British Journal of Psychiatry, 153*, 437-443.

Frith, C. D., Friston, K. J., Herold, S., Silbersweig, P., Fletcher, C., Cahill, R. J., Dolan, R., Frackowiak, S. J., & Liddle, P. F. (1995). Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task. *British Journal of Psychiatry, 167*, 343-349.

Gould, L. N. (1948). Verbal hallucinations and activity of vocal musculature; an electromyographic study. *American Journal of Psychiatry, 105*, 367-372.

Green, M. F., & Kinsbourne, M. (1989). Auditory hallucinations in schizophrenia: Does humming help? *Biol Psychiatry, 25*, 633-635.

Hoffman, R. (2010). Comment on vercammen, knegtering, den boer, liemburg, & aleman submitted 29 april, 2010. *Schizophrenia Research Forum*, Comment.

Hoffman, R. E. (1999). New methods for studying hallucinated 'voices' in schizophrenia. *Acta Psychiatrica Scandinavica Supplementum, 99*, 89-94.

Hoffman, R. E., Hawkins, K. A., Gueorguieva, R., Boutros, N. N., Rachid, F., Carroll, K., & Krystal, J. H. (2003). Transcranial magnetic stimulation of left temporoparietal cortex and medication-resistant auditory hallucinations. *Arch Gen Psychiatry, 60*, 49-56.

Hunter, M. D. (2004). Locating voices in space: A perceptual model for auditory hallucinations? *Cognitive Neuropsychiatry, 9*, 93-105.

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychon Bull Rev, 12*, 852-857.

Jardri, R., Delevoye-Turrell, Y., Lucas, B., Pins, D., Bulot, V., Delmaire, C., Thomas, P., Delion, P., & Goeb, J. L. (2009). Clinical practice of rtms reveals a functional dissociation between agency and hallucinations in schizophrenia. *Neuropsychologia, 47*, 132-138.

Jaspers, K. (1968/1912). The phenomenological approach in psychopathology. *British Journal of Psychiatry, 114*, 1313-1323.

Johns, L. C., & McGuire, P. K. (1999). Verbal self-monitoring and auditory hallucinations in schizophrenia. *Lancet, 353*, 469-470.

Johns, L. C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., & K, M. P. (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine, 31*, 705-715.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.

Kapur, S. (2003). Psychosis as a state of aberrant salience: A framework linking biology, phenomenology, and pharmacology in schizophrenia. *American Journal of Psychiatry, 160*, 13-23.

Keefe, R., Arnold, M., Bayen, U., McEvoy, J., & Wilson, W. (2002). Source-monitoring deficits for self-generated stimuli in schizophrenia: Multinomial modeling of data from three sources. *Schizophrenia Research, 57*, 51-67.

Keefe, R. S. E., Arnold, M. C., Bayen, U. J., & Harvey, P. D. (1999). Source monitoring deficits in patients with schizophrenia; a multinomial modelling analysis. *Psychological Medicine, 29*, 903-914.

Larøi, F., Van Der Linden, M., & Marczewski, P. (2004). The effects of emotional salience, cognitive effort and meta-cognitive beliefs on a reality monitoring task in hallucination-prone subjects. *British Journal of Clinical Psychology, 43*, 221-233.

Larøi, F., & Woodward, T. S. (2007). Hallucinations from a cognitive perspective. *Harvard Review of Psychiatry, 15*, 109-117.

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., David, A. S., Frackiowack, R. S. J., & Frith, C. D. (1995). Abnormal monitoring of inner speech: A physiological basis for auditory hallucinations. *The Lancet, 346*, 596-600.

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., Frackowiak, R. S., & Frith, C. D. (1996). The neural correlates of inner speech and auditory verbal imagery in schizophrenia: Relationship to auditory verbal hallucinations. *British Journal of Psychiatry, 169*, 148-159.

Menon, M., & Woodward, T. S. (2007). A comparison of one-high-threshold and two-high-threshold multinomial models of source monitoring. *Journal of Modern Applied Statistical Methods, 6*, 279-290.

Mintz, S., & Alpert, M. (1972). Imagery vividness, reality testing, and schizophrenic hallucinations. *Journal of Abnormal Psychology, 79*, 310-316.

Moritz, S., & Laroi, F. (2008). Differences and similarities in the sensory and cognitive signatures of voice-hearing, intrusions and thoughts. *Schizophr Res, 102*, 96-107.

Moritz, S., & Woodward, T. S. (2007a). Metacognitive training for schizophrenia patients (MCT): A pilot study on feasibility, treatment adherence, and subjective efficacy. *German Journal of Psychiatry, 10*, 69-78.

Moritz, S., & Woodward, T. S. (2007b). Metacognitive training in schizophrenia: From basic research to knowledge translation and intervention. *Current Opinion in Psychiatry, 20*, 619-625.

Morrison, A. P., & Haddock, G. (1997). Cognitive factors in source monitoring and auditory hallucinations. *Psychological Medicine, 27*, 669-679.

Morrison, A. P., Haddock, G., & Tarrier, N. (1995). Intrusive thoughts and auditory hallucinations: A cognitive approach. *Behavioural and Cognitive Psychotherapy*, 265-280.

Nayani, T. J., & David, A. S. (1996). The auditory hallucinations: A phenomenological survey. *Psychological Medicine, 26*, 117-189.

Nelson, N. W., Boone, K., Dueck, A., Wagener, L., Lu, P., & Grills, C. (2003). Relationships between eight measures of suspect effort. *Clin Neuropsychol, 17*, 263-272.

Northoff, G., & Qin, P. (2011). How can the brain's resting state activity generate hallucinations? A 'resting state hypothesis' of auditory verbal hallucinations. *Schizophrenia Research, 127*, 202-214.

Pinter, R. (1913). Inner speech during slient reading. *Psychological Review, 20*, 129-153.

Plaze, M., Paillere-Martinot, M. L., Penttila, J., Januel, D., de Beaurepaire, R., Bellivier, F., Andoh, J., Galinowski, A., Gallarda, T., Artiges, E., Olie, J. P., Mangin, J. F., Martinot, J. L., & Cachia, A. (2011). "Where do auditory hallucinations come from?"--a brain morphometry study of schizophrenia patients with inner or outer space hallucinations. *Schizophr Bull, 37*, 212-221.

Rankin, P. M., & O'Carroll, P. J. (1995). Reality discrimination, reality monitoring and disposition towards hallucination. *British Journal of Clinical Psychology, 34*, 517-528.

Rapin, L., Loevenbruck, H., Dohen, M., Metzak, P. D., Whitman, J. C., & Woodward, T. S. (in revision). Hyperactivity of voice-selective cortical regions in schizophrenia. *Psychiatry Research: Neuroimaging*.

Rotarska-Jagiela, A., van de Ven, V., Oertel-Knochel, V., Uhlhaas, P. J., Vogeley, K., & Linden, D. E. (2010). Resting-state functional network correlates of psychotic symptoms in schizophrenia. *Schizophr Res, 117*, 21-30.

Seal, M. L., Aleman, A., & McGuire, P. K. (2004). Compelling imagery, unanticipated speech and deceptive memory: Neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry, 9*, 43-72.

Seal, M. L., Crowe, S. F., & Cheung, P. (1997). Deficits in source monitoring in subjects with auditory hallucinations may be due to differences in verbal intelligence and verbal memory. *Cognitive Neuropsychiatry, 2*, 273-290.

Shergill, S., Bullmore, E., Simmons, A., Murray, R., & McGuire, P. (2000). Functional anatomy of auditory verbal imagery in schizophrenic patients with auditory hallucinations. *The American Journal of Psychiatry, 157*, 1691-1693.

Sommer, I. E., Diederen, K. M., Blom, J. D., Willems, A., Kushan, L., Slotema, K., Boks, M. P., Daalman, K., Hoek, H. W., Neggers, S. F., & Kahn, R. S. (2008). Auditory verbal hallucinations predominantly activate the right inferior frontal area. *Brain, 131*, 3169-3177.

Stephan, K. E., Friston, K. J., & Frith, C. D. (2009). Dysconnection in schizophrenia: From abnormal synaptic plasticity to failures of self-monitoring. *Schizophrenia Bulletin, 35*, 509-527.

Stephane, M., Kuskowski, M., McClannahan, K., Surerus, C., & Nelson, K. (2009). Evaluation of speech misattribution bias in schizophrenia. *Psychological Medicine*, 1-8.

Stephane, M., Thuras, P., Nasrallah, H., & Georgopoulos, A. P. (2003). The internal structure of the phenomenology of auditory verbal hallucinations. *Schizophr Res, 61*, 185-193.

Stone, J. M., Abel, K. M., Allin, M. P., van Haren, N., Matsumoto, K., McGuire, P. K., & Fu, C. H. (2011). Ketamine-induced disruption of verbal self-monitoring linked to superior temporal activation. *Pharmacopsychiatry, 44*, 33-48.

Suzuki, M., Yuasa, S., Minabe, Y., Murata, M., & Kurachi, M. (1993). Left superior temporal blood flow increases in schizophrenic and schizophreniform patients with auditory hallucination: A longitudinal case study using 123i-imp spect. *Eur Arch Psychiatry Clin Neurosci, 242*, 257-261.

Van der Gaag, M. (2006). A neuropsychiatric model of biological and psychological processes in the remission of delusions and auditory hallucinations. *Schizophrenia Bulletin, 32*, S113-S122.

Vercammen, A., Knegtering, H., Liemburg, E. J., den Boer, J. A., & Aleman, A. (2010). Functional connectivity of the temporo-parietal region in schizophrenia: Effects of rtms treatment of auditory hallucinations. *J Psychiatr Res, 44*, 725-731.

Waters, F., Woodward, T. S., Allen, P., Aleman, A., & Sommer, I. (2010). Self-recognition deficits in schizophrenia patients with auditory hallucinations: A meta-analysis of the literature. . *Schizophrenia Bulletin, DOI:10.1093/schbul/sbq144*.

Woodward, T. S., & Menon, M. (2010). Considerations for analysis of source monitoring data when investigating hallucinations in schizophrenia research. *European Archives of Psychiatry and Clinical Neuroscience, 261*, 157-164.

Woodward, T. S., Menon, M., & Whitman, J. C. (2007). Source monitoring biases associated with hallucinations in schizophrenia. *Cognitive Neuropsychiatry, 12*, 477-494.

Acknowledgements

TSW is supported by a Scholar award from the Michael Smith Foundation for Health Research (MSFHR) and a New Investigator award from the Canadian Institutes of Health Research (CIHR). The authors thank Paul Metzak and Frank Laroi for their contributions to this work.

Table 1. Conceptual model for the alienation and misattribution steps thought to underlie hallucinations. The Subjective Generating Origin dimension refers to the perceived/subjective origin of a given cognitive event, and the Subjective Source to the localization of the cognitive event in space. A combination of these two dimensions gives rise to four different types of cognitive events. Inner, self-generated (ISG) cognitive events are healthy and normal. All hallucinations originate as inner, self-generated cognitive events, but are somehow altered such that they are experienced as one of the other three cells in Table 1 (OSG, ONG, or ING). In this context, hallucinations may be viewed as inner, self-generated events that are first alienated and then misattributed, hence the reference to misattribution models (Larøi & Woodward, 2007).

	Subjective Generating Origin	
Subjective Source	Self	Nonself
Outer	Outer, self-generated (OSG)	Outer, nonself-generated (ONG)
Inner	***Inner, self-generated (ISG)***	Inner, nonself-generated (ING)

Table 2. Details regarding sample size and number of items per condition/source combination design for source monitoring studies where hallucinating patients were compared to non-hallucinating patients on externalization-type memory errors.

Authors	Groups (n)	Number of items/conditions	Notes
Bentall et al (1991)	Hallucinating schizophrenia patients (22) Non-hallucinating psychiatric controls (16) Healthy controls (22)	Self-generated (from cue) (16) Experimenter generated (from cue) (16) New (8)	Hallucinators made more externalization errors on high cognitive effort items
Brebion et al (2000)	Hallucinating schizophrenia patients (22) Non-hallucinating schizophrenia patients (18) Healthy controls (40)	Experimenter generated (8) Self-generated (from picture) (8) Self-generated (from category) (8) New distracters (24)	Hallucinators showed increased non-self misattribution errors
Brunelin et al. (2006)	Hallucinating schizophrenia patients (30) Non-hallucinating schizophrenia patients (31)	Self-generated (from cue) (8) Self-Imagined (from cue) (8) Experimenter generated (8) Experimenter imagined (8) New items (8)	Hallucinators showed increased externalization errors
Franck et al (2000)	Hallucinating schizophrenia patients (7) Non-hallucinating schizophrenia patients (5)	Self-internal (read silently) (8) Self-external (read out loud) (8) New items (8)	Hallucinators had a tendency to misattribute internal (i.e. read silently) and new words having been read out loud (p = .09)
Seal et al (1997)	Hallucinating schizophrenia patients (10) Non-hallucinating schizophrenia patients (12) Healthy controls (15)	Self-generated (from cue) (16) Experimenter generated (from cue) (16) New (16)	Hallucinators did not show source monitoring deficits compared to non-hallucinators when IQ was covaried
Woodward et al (2007)	Hallucinating schizophrenia patients (16) Non-hallucinating schizophrenia patients (35) Healthy controls (20)	Self generated (from cue) (25) Computer generated (25) Experimenter generated (25)	Externalization bias specific to hallucinations and not seen when group is split on delusional status.
Costafreda et al (2008)	Hallucinating schizophrenia patients (14) Delusional schizophrenia patients (6) Non-symptomatic schizophrenia patients (10)	Self-generated (from category) 24 Experimental generated (24)	Externalizations were more common in psychotic compared to non-symptomatic patients, and with negative stimuli.