Mind and Language 23 (2008): 369-401

Fractured Phenomenologies: Thought Insertion, Inner Speech, and the Puzzle of Extraneity

Peter Langland-Hassan

Abstract: How it is that one's own thoughts can seem to be someone else's? After noting some common missteps of other approaches to this puzzle, I develop a novel cognitive solution, drawing on and critiquing theories that understand inserted thoughts and auditory verbal hallucinations in schizophrenia as stemming from mismatches between predicted and actual sensory feedback. Considerable attention is paid to forging links between the first-person phenomenology of thought insertion and the posits (e.g. efference copy, corollary discharge) of current cognitive theories. I show how deficits in the subconscious mechanisms regulating inner speech may lead to a 'fractured phenomenology' responsible for the reports of inserted thoughts and auditory verbal hallucinations of schizophrenic patients. Supporting work on virtual environments is discussed, and lessons concerning the fixity of delusional belief are drawn.

Earlier versions of this work were presented at Toward a Science of Consciousness in Tucson, Arizona, the Center for Experimental Psychology's 'Exploring the Boundaries of Experience and Self' conference at Oxford, and at the CUNY Graduate Center's Cognitive Science Symposium. Many thanks to the questioners in the audiences at all those events. Special thanks also to Joëlle Proust, Galen Strawson, Shaun Gallagher, and two referees for this journal for their helpful comments and suggestions.

Address for correspondence: Department of Philosophy, The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016.
Email: plangland-hassan@gc.cuny.edu

1. Introduction

I will raise a question—called 'the puzzle of extraneity'—concerning the nature of auditory verbal hallucinations (AVHs) and inserted thoughts in schizophrenia. The puzzle of extraneity asks how it is that one's own thoughts can seem to be someone else's. A popular proposal (Campbell, 1999; Gallagher, 2000; Graham and Stephens, 2000; Proust, 2006) has been that the normally fused subjective senses of thought *ownership* and of thought *agency* come apart in schizophrenia. It is the missing 'sense of agency' and lingering 'sense of ownership' (the latter deriving from the introspective mode of access to the thought) that is usually thought to account for an introspected thought's being ascribed to another (Campbell, 2002). Yet there is little agreement concerning what the phenomenological 'sense of thought agency' amounts to. One goal of this paper is to show that the phenomenology of thought agency is not what one might pre-reflectively take it to be, and to provide a way of understanding this phenomenology that gives genuine purchase on the puzzle of extraneity.

After rejecting two competing solutions to the puzzle (in section three), I develop a novel solution (sections four and five), drawing on, and amending, the work of Frith (1992, 2005), Campbell (1999), Blakemore (2003, 2005), Jeannerod and Pacherie (2004) and Proust (2006). The central argument of section four is that the popular explanatory paradigm—originating with Frith (1992)—invoking 'comparator' mechanisms responsible for calculating the relationship between predicted and actual perceptual (and cognitive) feedback can be understood in at least two different ways. I discuss these differences, favoring what I call the 'filter model' of sensory cancellation. In section five I examine in some detail the potential consequences that deficits in such filtering would have for the phenomenology of thought (and, in particular, of inner speech). Specifically, I try to indicate why such phenomenology would lead to reports both of inserted thoughts and hallucinated 'voices'. This proposal is supplemented (in section seven) by a discussion of recent research concerning the factors needed to

create 'presence'[1] in a virtual environment (VE) (Sanchez-Vives and Slater, 2005). VE research pinpoints some of the cognitive mechanisms underlying the fixity (i.e., the resistance to revision in light of contravening evidence) of delusional belief in a way that supports the present hypothesis, and points the way towards its further testing.

2. The 'Dual Strands' of Thought Ownership

Auditory verbal hallucinations and inserted thoughts are two cardinal (so-called 'first-rank' (Schneider, 1959)) symptoms of schizophrenia. AVHs are diagnosed when patients report hearing voices in the absence of appropriate stimuli (whether or not the patients believe themselves to be merely hallucinating). Inserted thoughts—typically classified as delusions and not hallucinations—are diagnosed when patients report feeling as though thoughts have been inserted into their minds by a foreign agent. Patients experiencing thought insertion do not take themselves to be hearing the voices of others, but rather feel as though they are subjectively entertaining the thoughts of another person (or even, that someone else is *using* their thoughts to think). Surveying 100 schizophrenic patients prone to AVHs, Nayani and David (1996) found that forty-six reported experiencing the distinct phenomenon of thought insertion, where patients' thoughts were reported as 'lacking normal sense of ownership,' seeming 'alien, not their own' (p. 182).

Inserted thoughts, like AVHs, often take the form of verbal comments or commands. Frith (1992, p. 66) quotes one patient's account:

> Thoughts are put into my mind like 'Kill God.' It is just like my mind working, but it isn't. They come from this chap, Chris. They are his thoughts.

Another commonly cited example comes from Mellor (1970, p. 17):

[1] 'Presence' is a technical term in the VE (virtual environment) literature meant to capture a VE's ability to instill in the perceiver a sense of actually 'being there' in the VE (Sanchez-Vives and Slater, 2005).

> I look out the window and I think that the garden looks nice and the grass looks cool, but the thoughts of Eamonn Andrews come into my mind. There are no other thoughts there, only his….He treats my mind like a screen and flashes thoughts onto it like you flash a picture.

Notice that in each case the patient identifies the 'inserted' thoughts as being located in 'my mind'. This leads to a philosophical puzzle: how can someone acknowledge that a thought occurs in her mind while denying that it is her thought? A seemingly inviolable relation is being flaunted: isn't any thought in my mind *ipso facto* my own thought? A powerful intuition holds that, when it comes to thoughts, *location determines ownership* (unlike laptops, for instance).

As already remarked, a common way of making sense of such reports is to draw a distinction between the subjective senses of thought ownership and of thought agency, and to suggest that these normally fused aspects of phenomenology come apart in sufferers of inserted thoughts (Campbell, 2002, p. 36). To have a subjective sense of thought ownership without a sense of thought agency is, intuitively, to experience a kind of immediate, (apparently) non-inferential (or 'private') *access* to the content of a thought, while at the same time lacking any phenomenological sense of being causally responsible for the generation of that thought (Campbell, 1999).

Once we unpack the idea in this way, however, it becomes evident that normal subjects also experience this 'split' from time to time, such as when one finds oneself with a song fragment repeating in one's mind, or simply engages mechanistically in a routine activity (an activity which nevertheless involves conscious thought). These quotidian experiences do not lead us to ascribe the agency or ownership of such thoughts to an external agent. One might also question whether most thought really is accompanied by a sense of causal agency or 'effort' at all (Strawson, 2003). Perhaps when solving complex problems or negotiating new situations, we may have a robust sense of being the causal agents of our thoughts and decisions—of controlling and shaping the course of our thinking—however it is far from clear that at other times there is

anything properly called a 'phenomenology of agency' attending thought (at least, not if we think of this phenomenology as an effortful sense one has of *trying* to think in certain ways). The unsupported assumption that there is a clearly understood phenomenology or 'emotion' of thought agency—one that can explain reports of thought insertion by its absence—is a pervasive weakness in the literature on this subject.[2]

Thus, simply appealing to a dissociation between 'two strands' in the ordinary conception of thought does not take us very far in answering the puzzle of extraneity. Two challenges remain: first, a clear account of the nature and causal origin of the agency-phenomenology that becomes abnormal in thought insertion must be given. This account should be consistent with the fact that we often think without experiencing any sense of effort or of an exertion of cognitive force. Second, the phenomenology of thought insertion should be explained in a way that reveals why schizophrenic patients not only fail to feel causally responsible for their thoughts, but also feel that their thoughts *belong to someone else* (thereby explaining the *extraneity* attending their phenomenology). To meet these challenges is to solve the puzzle of extraneity.[3]

3. Alternative proposals

Before going on to develop a positive solution to the puzzle of extraneity, it will help to consider some rival accounts—both in order to appreciate the depth of the puzzle, and to reveal a connection between auditory hallucinations and inserted thoughts that will be important in later sections.

---

[2] Cf. Frith (1992, p. 81), and, more recently, Jones and Fernyhough (2006) and Proust (2006).

[3] While the extreme oddity of schizophrenic symptoms may tempt one in the direction of skepticism concerning psychological, phenomenological, or cognitive explanations (it may seem that there is no longer a stable enough belief-desire economy in schizophrenia to ground a cognitive theory), this outlook is unsupported by the empirical literature. Maher (1974, p. 99) cites a wide range of studies indicating that 'when adequate controls for variables such as education, verbal knowledge, etc. are introduced, there is no basis in evidence for the belief that schizophrenic patients differ from normals in their vulnerability to any type of [logical] error.' And, indeed, many of those afflicted with schizophrenia carry on successfully in day-to-day reasoning tasks, the peculiarities of their experience revealed mainly in the reports of their hallucinated voices or inserted thoughts, the content of which is typically 'fairly narrow' (Frith, 1992, p. 79).

3.1 The Sensory Resemblance Thesis

Perhaps the first idea that comes to mind in trying to answer the puzzle of extraneity is that patients may ascribe the agency or ownership of both auditory verbal hallucinations and inserted thoughts to others simply because what it is like to hear voices and have inserted thoughts is similar to what it is like to actually hear others speak—in essence, like causes are inferred from like effects (Slade and Bentall, 1988; Kinsbourne, 1990). Call this the Sensory Resemblance Thesis (SRT). SRT may indeed solve the puzzle of extraneity for some auditory hallucinations. However, closer attention to the reported phenomenology of auditory hallucinations and inserted thoughts shows SRT to be an insufficient solution to the puzzle of extraneity for most such episodes. To see why, we have to look more closely at the relationship between AVHs and inserted thoughts.

Given that inserted thoughts can take the form of verbal commands and commentary, and are experienced as unwilled by the subject, how are they different, phenomenologically, from auditory hallucinations? One natural suggestion is that auditory hallucinations are simply more vivid, or posses richer sensory character, making them more likely to be mistaken for cases of actually hearing someone speak, resembling less the mere verbal-imagery-involving thought (or 'inner speech') that most people engage in from time to time (as when one debates an imagined opponent in an inner dialog). However, some patients report experiencing *voices* (as opposed to inserted thoughts) that lack any auditory or sensory component (Graham and Stephens, 2000, p. 99-100). How should *these* experiences be distinguished from thought insertion?

The correct answer may be that they should *not* be so distinguished. Graham and Stephens take this approach, giving reasons for thinking that auditory hallucinations and inserted thoughts should be treated as kindred phenomena, the distinction arising partly out of patients' difficulties in reporting the nature of their experiences, and partly out of differences in the severity of a single disorder. In support of this idea, they discuss the work of Junginger and

Frame (1985), who asked 26 hallucinating patients to rate their voices on a 1-10 scale, where 1 indicated that the voice seemed to exist entirely 'inside the head' and 10 indicated the voice seemed clearly to come from 'outside the head' (i.e., was like hearing someone else speak). Their results were mixed: 11 subjects (41%) rated their voices at 5 or below, while 15 (59%) rated them at 6 to 10. A minority of patients reported that hearing voices was more like 'hearing yourself think' than like hearing somebody talk. Nayani and David's (1996) survey supports these findings: 49% of patients reported hearing voices 'through their ears as external stimuli', 38% reported hearing them 'in internal space,' and 12% heard them in both loci variably (p. 180).

Such reports weigh in favor of two conclusions: first, many episodes classified as auditory hallucinations (and all inserted thoughts) are not, for their subjects, similar to experiences of actually hearing someone speak (normally perceived voices do not seem to come from inside the head, or to lack sensory character). Therefore, the agency responsible for such episodes must be outwardly ascribed for some reason other than what the SRT posits.
Second, we have reason to think that some reports of voices (those towards the lower end of Junginger and Frame's scale) may be reports of the very same phenomenon that others report as inserted thoughts. Just as one might, in forming a visual image of one's childhood home, say 'I can see it now,' (even if having visual imagery is only a little like seeing), so might one who experiences unusual inner speech say that he 'hears' various comments and commands. In short, patients may be using normal modes of expression in a rough-and-ready way, in order to communicate the nature of experiences that do not fall neatly into any preexisting category.

These reflections are supported by the pioneering work of Eugen Bleuler (1911/1950, p.110), who found that his schizophrenic patients sometimes described their voices as 'soundless' and as like 'vivid thoughts.' Bleuler notes that 'even intelligent patients are not always sure that they are actually hearing the voices or whether they are only compelled to think them' (*ibid.*). Though Bleuler's work predates the auditory hallucination/thought insertion distinction,

something very much like it is present in his distinction between 'hallucinations of perceptions' and 'hallucinations of conceptions.' These are 'differentiated by patients' accordingly as the voices appear to 'come from the outside' (hallucinations of perceptions) or, alternatively, seem 'projected into their own bodies' having 'hardly any sensory components' (hallucinations of conceptions) (1911/1950, p. 111). Here, too, the key role played by the patient in 'differentiating' one phenomenon from the other should encourage us to leave open the possibility that the very distinction arises out of alternate ways of describing a single unusual phenomenon. Also, given that the mechanisms responsible for more 'vivid' auditory hallucinations likely overlap with those responsible for inserted thoughts and 'soundless' voices, we should also leave open the possibility that whatever solves the puzzle of extraneity for the latter may also solve it for the former.

3.2 Graham and Stephens' 'Intentionally Inexplicable' Thoughts

Graham and Stephens take a different approach to solving the puzzle of extraneity, arguing that it is the 'apparent intelligence' of inserted thoughts and hallucinated voices that provides the 'experiential or epistemic basis' for the attribution of their agency to another (2000, p. 174). Of course, for the apparent intelligence of a thought to lead one to other-ascribe its agency, one must *first* not take oneself to be the agent of that thought (unless one is utterly convinced of one's own unintelligence). According to Graham and Stephens, a person's initial *lack* of a sense of agency results from a given thought or mental episode not being 'explicable in terms of his conception of what he believes and desires' (2000, p. 165). On their view, people continually monitor the content of their own thoughts and, when thoughts are detected which do not 'fit' their other beliefs and desires (being, in Graham and Stephens' term, 'intentionally inexplicable'), these thoughts acquire an alien phenomenology. Thus, for Graham and Stephens, there is nothing *phenomenologically* foreign about inserted thoughts or hallucinated voices *prior to* their detection as being out of keeping with one's

beliefs and desires, allowing them to *explain* the phenomenology of non-self-agency in non-phenomenological terms.

If these ill-fitting thoughts or voices then *also* seem intelligently structured, an external ascription of the agency responsible for them will occur as a sort of inference to best explanation. In essence a kind of 'argument from design' is applied, roughly as follows:

P1) These thoughts I am entertaining require an active intelligence for their creation (they are not mere 'parrotings' or repetitions of earlier perceived events).
P2) I am not the active intelligence responsible for them (due to their lack of 'fit' with the self-narrative).
C) Someone else is creating/thinking these thoughts.

Often, the voices and inserted thoughts experienced by patients are accusatory and belligerent. Such voices and thoughts—apparently expressive of an antagonistic 'intelligence' of some kind—clearly oppose one's larger goals and self-conception, and provide the key examples for Graham and Stephens' account of the sense of thought-agency. However, as Graham and Stephens themselves note, many voices and inserted thoughts are generally positive in their content, their messages consistent with the larger plans and desires of their subjects. Bleuler notes that 'besides their persecutors, the patients often hear the voice of some protector' (1911/1950, p. 98). He also reports that, occasionally, hallucinatory voices 'represent sound criticism of [the patient's] delusional thoughts and pathological drives' and therefore should be generally consistent with whatever non-delusional thoughts he may still have (*ibid.*). Graham and Stevens also discuss Chapman and Chapman's (1998) report of a young woman who persistently heard the voices of her deceased father and grandmother. She regarded these voices as benevolent and generally beneficial to the achievement of her practical goals (such as when they advised her on the purchase of a new car) (2000, p. 169-170).[4] Such cases show that one's inability

[4] In light of such cases, Graham and Stephens suggest 'it may be necessary to tell different stories about why the subject finds her thoughts intentionally inexplicable in various cases' (2000, p. 171). It is not clear that this explanatory blank check can be cashed. If showing a thought to

to intentionally explain a thought is not necessary in order for that thought to have an element of extraneity.

Nor is one's inability to intentionally explain a series of intelligently structured thoughts sufficient to make them seem to be owned or produced by another agent. Few people, before the onset of mental illness, believe in the existence of (actually) inserted thoughts and (actually) perceived voices of invisible agents, so any intelligently structured belief in such things will itself (at least initially) be radically out of keeping with one's other beliefs and desires. Graham and Stephens' theory should predict that the very thought that one has had thoughts inserted into one's mind (or that one is hearing the voice of an invisible agent) will itself seem not to be one's own thought. And yet these (delusional) thoughts are precisely the ones that *are* self-ascribed by schizophrenic patients, and are the ones theorists must explain. It is plainly self-defeating for a subject to other-ascribe the agency of her own thoughts for the sole purpose of preserving a consistent self-narrative, for such a claim immediately turns one into the protagonist of a very different sort of story.

It is worth noting that Graham and Stephens' theory may still explain the lack of a sense of thought-agency that attends certain *non-pathological* thought episodes. Take, for instance, the novelist who feels she must write in order to discover what her characters will do, or the song-writer who feels he is merely a medium through which some other agency is expressed. Unable to gain conscious access to the aspects of their psychologies responsible for their artistic creations (i.e., unable to intentionally explain their acts), artists—and highly-trained individuals in many other fields—occasionally feel as though they are not the agents of their thoughts and actions. However, this mere lack of an intellectual understanding of how one accomplishes one's acts is a far cry from the pathological phenomenology of other-agency that characterizes schizophrenic experience. The obvious difference remains that this 'artist's extraneity' rarely reaches the point where a delusional belief is formed that

---

be intentionally inexplicable requires considerable explanation from case to case, the overall form of explanation will inevitably seem *ad hoc*.

someone else actually is the agent of one's thoughts. This is a difference that clearly calls for a different explanation in terms of the cognitive mechanisms responsible for the respective kinds of cases. Nevertheless, it is helpful to have Graham and Stephens' account at hand to explain cases where superficially similar reports to those of schizophrenic patients are made by normal subjects, if only to more clearly distinguish them from the cases of thought insertion to be discussed.

To briefly consider a further option, one can simply posit a certain phenomenological 'my-ness' (perhaps taken for granted in normal thinking) which is lacking in the case of inserted thoughts and 'soundless' auditory hallucinations, a palpable 'otherness' existing in its stead. Yet, this sort of phenomenological 'answer' to the puzzle of extraneity does little more than restate the problem and assure us that the patients are not lying—it tells us that *it really does seem to them* that their thoughts are controlled (or owned) by another. Though doubtless true, this does nothing to explain *what it is* for their thoughts to lack the normal phenomenological character of 'my-ness'. Just what is this 'my-ness', other than a label for the problem? And why think that we can easily form a clear notion of what it would be like for it to go missing? Those who have never had such experiences may be no better suited to imagining them than the blind are to imagining sight. So, the question of the *character* of the delusional experience remains open.

Nor do phenomenological answers explain *how* such phenomenology could come to exist in an otherwise rational being. We still need to know which mechanisms, present in normal humans, can malfunction to lead to these specific symptoms. To avoid seeming *ad hoc*, or a mere restatement of the problem, the mechanisms cited in the explanation should serve a role in the explanation of phenomena outside of the deficits they are here called on to explain. A last criticism of purely phenomenological approaches concerns their appeal to normally unnoticed or 'overlooked' elements of phenomenology (the 'sense of agency' or 'my-ness' of thoughts). Given that a creature's phenomenology is just a matter of how things seem to that creature, the idea of a

'taken for granted' or unnoticed aspect of phenomenology is obviously problematic.

4. Sensory-Feedback Approaches

While phenomenological approaches face problems, purely neurological approaches have shortcomings as well. It has been known since the late 1950s that schizophrenic symptoms are correlated with increased levels of dopamine absorption in the brain (indeed, this is one of the main reasons researchers are confident the disease has a unitary biological basis)—hence the common prescription of dopamine-blockers as antipsychotic drug treatments. However, the theoretical vocabulary of 'dopamine' and 'neurotransmitters', etc., is too distant in terms of causal/theoretical links from that of psychology to provide the kind of explanatory purchase on schizophrenia that would allow for a more nuanced (and perhaps even curative) treatment. For all current researchers know, increased dopamine absorption may be an effect of the underlying disorder, and not its cause. Hence the move theoretically 'upward' to the cognitive-modular level of explanation.[5]

At this level of explanation, the oddity of schizophrenic reports is attributed to correspondingly unusual (and highly variable) experiences. These may be experiences normal perceivers do not know what it is like to have, yet which nevertheless arise from commonly shared cognitive mechanisms and processes. Importantly, the mechanisms and processes appealed to in such explanations are ones that have a theoretical life outside of the peculiar puzzle they are here called on to explain.

4.1 Frith, Efference Copies, and the Attenuation of Sensory Input

Frith (1992) was one of the first to develop a detailed cognitive analysis of this kind—a 'box and arrow' analysis pitched at the level of functionally specialized

[5] For an extended discussion of the issue of levels of explanation in schizophrenia by a cognitive scientist, see Frith (1992, p. 25-29).

cognitive mechanisms, which map (very) roughly onto various regions of the brain. The operation of such mechanisms often involves processes and informational (or representational) states that are themselves not available to consciousness; their existence is typically posited based on comparative studies in psychology and neurology. In essence, a level of explanation is sought which may provide more explanatory resources than those available at the commonsense phenomenological level, but which is also functionally *near enough* the level of psychological explanation for its (mal)functioning to have clear implications for the specific kinds of abnormalities reported by schizophrenic patients.

A guiding insight of Frith's (1992, 2005) account of the cognitive basis of schizophrenic delusions and hallucinations is that the prediction (and subsequent monitoring) of the sensory consequences of an action plays a key role in allowing the subject to distinguish between self and other-generated changes in sensory input. This link between perception and the phenomenology of agency—accepted also by Blakemore (2005) and Jeannerod and Pacherie (2004), among others—stems from the recognition of a problem faced by the early perceptual system. Moving one's eyes causes changes in the signal detected by the retina, but there is nothing in this changing signal *per se* that allows the perceptual system to determine whether the change is caused by one's own willful eye and head movements, or by an independent event perceived in the environment. Sperry's (1950) (and, concurrently, von Holst and Mittelstaedt's (1950/1973)) well-known solution to this problem (drawing on earlier work by Helmholtz (1866)) was to propose that motor commands resulting in eye movements release an 'excitation pattern' that includes a signal (often called 'efference copy') which tells the perceptual system to anticipate certain kinds of changes in incoming retinal information (the incoming sensory information is called 'reafference'). Sperry calls this 'adjustment factor' allowing for the anticipation of certain kinds of input a 'corollary discharge', and suggests it could 'aid in maintaining stability of the visual field under normal conditions during the onset of sudden eye, head, and body movement' (Sperry, 1950, p.488).

A simple experiment reveals the power of this proposal to shed light on the phenomenology of agency. With one eye closed (or a hand over one eye), push lightly several times on side of the eyelid of your open eye. Your visual image should appear to 'move' or 'wobble'. However, if one creates a similar retinal change by simply moving one's head in the proper direction, no such 'wobbling' of the visual image occurs.[6] Why?

Sperry's explanation is that applying pressure to the eyeball with one's finger results in changes of the retinal image in the absence of the normally triggered efference copy signal (and resultant 'corollary discharge').[7] The efference copy signal is not triggered in this case because one's retinal display does not typically change as a result of a finger pressing against the eye. However, normal eye, head, and body movements *do* typically result in retinal changes, and therefore are accompanied by the firing of efference copy signals. The corollary discharge triggered by these signals filters out potentially distracting or irrelevant elements of the incoming information so as to result in a stabilized visual image. Without the stabilizing effect of the corollary discharge, the visual phenomenology is as of the room itself moving—and, indeed, if it were not for contrasting information simultaneously registered by other sensory and proprioceptive systems (and the knowledge that one has simply pushed one's eye), it would indeed seem as though the environment itself were moving (as in an earthquake), and not simply one's visual field. Thus, the same structure of changes on the retina can result in very different phenomenologies, depending on whether the efference copy/corollary discharge mechanisms are active. One phenomenology (the stable visual field) will naturally be associated with one's

[6] A similar phenomenon in present in the opposite direction as well: when eye muscles are partially paralyzed with curare (Brindley and Merton, 1960) and the subject told to move his eyes, the subject reports the visual appearance as of the room moving in the direction in which the eye movement would have occurred; though the retinal display has remained *exactly* the same, the entire room seems visually to shift in the direction of the merely intended movement, since, under normal conditions, that sort of retinal display, when filtered in accordance with an efference copy signal, would correlate with the room's actually moving in the direction of the shifted gaze.

[7] Some theorists (Gallagher, 2000; Campbell, 1999) identify corollary discharge with efference copy, while others (Campbell and Pettigrew, 2004) insist that the two are distinct phenomena—see fn. 8 for elaboration on this point.

own willful actions, while the other will be associated with perceptual changes caused by events outside of one's agency.

This 'filtering' of retinal information through the efference copy/corollary discharge process is just one instance of a general phenomenon, whereby the sensory consequences of self-initiated actions are 'automatically' (i.e. pre-consciously) attenuated (Blakemore, 2005). Studies by Blakemore *et al.* (1998, 2003), Ford and Mathalon (2005), Shergill *et al.* (2005), and Lindner *et al.* (2005) indicate that schizophrenic patients are often unable to achieve this attenuation, across multiple sense modalities (including proprioception). For instance, normal subjects are generally unable to tickle themselves, their sensitivity to their own touch being greatly attenuated (Weiskrantz, Elliot, and Darlington, 1971). Frith (2005) notes that this attenuation of self-generated sensation can also be observed at the physiological level, as activity in the somatosensory cortex is much reduced when tactile stimulation is self-applied. Schizophrenic patients do not show this attenuation, rating their sensitivity to self-touch as high as the same touch applied by someone else (Blakemore *et al.*, 1998). Also, brain imaging reveals that responses to sound in auditory cortex are normally attenuated when a person is speaking. This attenuation is not found in many schizophrenic patients (especially those who report auditory hallucinations) (Ford and Mathalon, 2005). Ford and Mathalon (2004) also found that, while inner speech reduces responsiveness of auditory cortex in normal subjects, schizophrenic patients lacked this attenuation during inner speech. In the case of vision, Lindner *et al.* (2005) found that schizophrenic patients with delusions of influence were more likely to attribute the sensory consequences of their own eye movements to the environment rather than to themselves.

4.2. Two Ways of Understanding Corollary Discharge

A potentially powerful explanatory paradigm invoking efference copies and reafferences—sometimes called the 'central monitoring' or 'comparator' theory—has been developed and adopted in one form or another by Frith (1992, 2005), Blakemore *et al.* (2002, 2003), Shergill *et al.* (2005), Linder *et al.* (2005),

Jeannerod and Pacherie (2004)[8], Campbell (1999), and others. These theorists all make explicit appeal to a 'comparison' that is made between an efference-copy-elicited 'prediction' (or 'corollary discharge') and reafferent (sensory or cognitive) input. Insofar as the two signals 'match', the incoming signal is said to be 'cancelled out' (resulting in its overall attenuation) (Blakemore *et al.*, 2002, p. 238).[9] However, this notion of a modality-specific *comparison* is not explicitly present in Sperry or von Holst's classical appeal to efference-copy-driven attenuations and cancellations of sensory feedback; rather, it grows mainly out of the work of Miall *et al.* (1993) and Wolpert *et al.* (1998) on 'forward models', hypothetically used by the motor system to adjust and control limb movements.

In brief, a 'forward model' of the kind posited by Miall *et al.* and Wolpert *et al.*—and expanded on in important ways by Grush (2004)—generates predictions of sensory feedback, based on efference copies of commands issued by the motor system. Some such predictive mechanisms seem necessary to explain corrective arm adjustments made in grasping tasks that occur too quickly (200-300ms) to result from the visual or proprioceptive monitoring of sensory feedback (Miall *et al.*, 1993, p. 205; 1998, p. 343; Wolpert and Ghahramani, 2000). If motor commands to engage in arm movement involve the triggering of an efference copy of the command, which itself generates a 'prediction' of the final state (perhaps by enacting an 'offline' emulation of the action (Grush, 2004)), this prediction can instantaneously be 'compared' with the achieved state in order to determine whether the desired movement was accomplished. Any discrepancies

---

[8] Jeannerod and Pacherie develop a simulation-based theory importantly different from Frith's theory in how it treats 'covert' actions, but nevertheless claim their theory to be 'largely complementary' to Frith's paradigm (2004, p. 125).

[9] There is no common understanding in the literature of what 'efference copy' and 'corollary discharge' refer to. Campbell (1999) and Grush (2004) identify efference copy with corollary discharge, while Blakemore *et al.* (2003, see esp. Fig. 1, p. 1059)—drawing on Miall *et al.* (1993)—use 'efference copy' to refer to the copy of the motor command that is sent to a later system that in turn generates a prediction of the incoming state, this prediction being called a 'corollary discharge'. Thomas Campbell and John Pettigrew (2004) agree that there is a distinction between efference copy and corollary discharge, but use the terms in precisely the opposite manner, suggesting that efference copy is the sensory prediction compared with input, and corollary discharge the copy of the motor command that results in the generation of the prediction.
I adopt Blakemore's *et al.* (2003) use: efference copy is a parallel 'copy' of a motor command, while corollary discharge is a prediction (or, in my case, a filter command) of the sensory consequences of the action, made based on the (efference) copy of the motor command.

between the forward output model's prediction and reafferent feedback detected by the comparator mechanism will cue the organism to adjust its behavior appropriately. If, however, the prediction and feedback 'match', the reafferent signal is cancelled out, making us 'largely unaware of the sensory feedback about the actual state of our motor system as long as our intentions have been achieved' (Blakemore *et al.*, 2002, p. 238). Notably, Miall *et al.* (1993), Wolpert *et al.* (1998), and Wolpert and Ghahramani (2000) only explicitly discuss and defend the operation of such models with regard to *limb movements,* and do not link their work to the classical models of Sperry and von Holst (Wolpert and Ghahramani (2000, p. 1217) do *anticipate*, however, that forward models 'will be found to underlie the control of motor systems as diverse as the eye, arm, speech, posture, balance and locomotion.').

Appropriately, Blakemore *et al.* (2003) draw on Miall *et al*. (1993) and Wolpert *et al.* (1998) to explain cases where schizophrenic patients do not feel in control of their own bodily movements. Yet, in attempting to extend such findings to explain auditory hallucinations and inserted thoughts, Frith (1992, 2005), Blakemore *et al.* (2003), Jones and Fernyhough (2006), and Blakemore (2005) run together discussions of Miall/Wolpert-type forward output models with Sperry and von Holst's more general account of corollary discharge, giving the impression that forward output models are the only (or at least best) way of implementing Sperry and von Holst's theories for the auditory and visual systems.

But there is another way of understanding how corollary discharge and sensory cancellation may take place in the visual and auditory systems—one which has some advantages when it comes to explaining inserted thoughts and AVHs. Before offering that account, I will describe in some more detail how the sort of forward output model typically discussed in the literature would apply to the visual system, so that the contrast between it and the theory I will offer is clear.[10]

[10] As I note below, some (but not all) accounts of forward output models (e.g., Wolpert and Ghahramani (2002)) are suitably abstract as to be neutral between the two accounts of sensory cancellation I here describe.

Suppose, for simplicity, that the retina is a two dimensional square grid containing nine cells, each of which can take a certain range of values. Again, for simplicity, assume the only possible values of these cells are ON and OFF. We can refer to the cells by their coordinates—the lowest left cell being (1,1), the upper right cell being (3,3), and so on. The present state of the retina can thus be given by a list indicating whether each coordinate pair is ON or OFF. According to most versions of the forward output model, the motor and/or visual system uses an efference copy of a motor command (e.g., look to the right) to calculate a prediction of the sensory feedback that will be received. This calculation requires as input both the command that is sent and an initial state of the retina (in the case of vision). Suppose that the nine-cell "retina" in its initial state (S1) has a value of ON for cells (2,1), (2,2) and (2,3), and of OFF for the rest. When an efference copy of the command "look to the right" is issued, a prediction state (P1) is generated, itself consisting of a nine-celled grid, where (let us suppose) the values are ON for cells (1,1), (1,2) and (1,3) and OFF for the rest.[11] P1 is then 'compared' with actual sensory input (S2). If S2 is in fact ON (1,1), (1,2) and (1,3) and OFF for the rest, then the prediction and subsequent input match, and the incoming information is cancelled (i.e. prevented from carrying forward to further processing). Note that, for this cancellation to occur, a separate 'comparator' mechanism (or algorithm) is needed that, in effect, makes the judgment that a match has (or has not) occurred between P1 and S2.

This is, I take it, an extremely simplified version how a forward output model involving predictions and comparisons could function in the visual system (it is, in essence, what Grush (2004, p. 189) proposes). Yet it seems that the same sensory cancellation could occur in a simpler way. We know that information on the retina changes in law-like ways each time one looks to the right. This means that, given an initial state of the retina, and a copy of a motor command to look to the right, the visual system could simply predict the needed cancellation without ever generating a simulation of the input itself (*ala* P1 above)

[11] Note that prediction state (P1) could potentially play its role while being a coarser grained representation than the state to which it is compared—i.e., while having a determinate value for only *some* coordinate pairs. Nothing in my discussion will hinge on this point.

to be compared with sensory input. That is, given S1 (defined as ON for (2,1), (2,2) and (2,3) and off for the rest), together with a command to look to the right, a well-trained (or adapted) visual system could simply issue another command to turn OFF cells (1,1), (1,2) and (1,3) in S2. In effect, for *every* case of looking to the right, the system could apply to each subsequent state the rule 'turn OFF each cell to the left of any cell that is ON in present state,' using each present state to determine which specific cells those should be from case to case.

Mathematically, this model accomplishes the same cancellation as the forward output model described above. It also can be said to feature a kind of prediction—the negative, input-cancelling state that is issued can be seen as an anticipation that a certain kind of eye movement will occur (and, therefore, that a specific kind of attenuation is needed). Yet there are important differences in how predictions and cancellations are accomplished by the two models. On the model I am suggesting (the 'filter model'), there is no need for a mechanism that *compares* the predictive signal to sensory input. The filter signal may simply be *applied* to S2 without need for a judgment to take place concerning its relation to S2, just as a sequence of delete commands could remove every third frame of a film without requiring any comparison between the deletion signals themselves and the film.

To briefly expand on this last metaphor, suppose that in the editing of a film a particular temporal sequence of delete commands (D1) is needed following any frame containing only red (because certain patterns of unnecessary frames invariably follow completely red frames). The triggering of D1 could be thought of as a prediction or anticipation that these unnecessary frames will be occurring. But there are not, literally, any states in the editing system or film that are *compared* with the deleted frames during this 'attenuation'.

Removing the need for a comparison allows the system to apply one less 'rule' during perception. Recall that, on the comparator or forward output model described above, the cognitive system takes S1 and a motor command as input, and applies a rule (an algorithm) to generate P1 (the prediction). Then an additional rule must be applied that governs how to change or attenuate S2 given

P1 (this is the rule or algorithm that, in effect, 'compares' P1 to S2). On the filter model, the cognitive system takes S1 and a motor command as input and simply applies an appropriate negative algorithm to S2.[12]

Note also that the filter model does not require that the 'predictive' signal itself be a quasi-visual state similar in kind to S1 and S2. The 'prediction' could simply be a negative signal that, like the delete command of the film editing system, has the effect of turning off certain targeted cells of S2, without itself representing the presence or absence of an object at some point in space (this is because it acts only to attenuate or delete certain parts of S2, while 'saying' nothing one way or another about the rest of the visual grid). The point, in more intuitive terms, is that we need not form an image in order to attenuate or 'filter' one.

Defenders of comparator or forward output models may reply that their models require nothing more than the kind of filtering I have suggested, and that the cancellation I have described can still be thought of as a sort of comparison. I find it odd to think of the described filtering as involving comparison, but the commonsense terminology used to describe such processing is ultimately unimportant. What I hope to have made clear is that there are (at minimum) two importantly different ways that sensory cancellation may occur. If both are ultimately consistent with a suitably abstract understanding of the forward output model paradigm, then at least some clarity concerning possibilities for its implementation has been achieved.

Many discussions of corollary discharge are simply too abstract to make it obvious which of the two described pictures is favored.[13] Some, however, clearly favor the initial picture, involving a comparison between a P1 state in expressed

[12] It is assumed on both models that, given S1 and a copy of the motor command, the cognitive system has the resources to generate an appropriate 'predictive' signal—be it a negative 'filtering' signal or a positive one sent to a comparator. How the system initially learns to generate this appropriate signal is an interesting further question—see Wolpert and Gharamani (2000, p. 1216-1217) for discussion.

[13] For instance, Ford and Mathalon (2004, p. 37) note that 'efference copy works to suppress perception when it results from self-generated action,' allowing 'an automatic distinction between internally-generated and externally-generated percepts.' They go on to describe a range of experiments implicating corollary discharge dysfunction in schizophrenic patients, while saying nothing that would favor one of the models I have discussed over the other.

visual coordinates and S2. Grush is explicit concerning the quasi-perceptual nature of the predictive states triggered by efference copies, arguing that 'emulators' (his term for 'forward models') take motor commands as input and produce 'an output signal…identical or similar to the feedback signal produced by the plant'[14] (2004, p. 379). On his theory, visual imagery itself is '"mock" input generated from the operation of an internal emulator' offline (2004, p. 390). A commitment to predictive and input states being of the same essential kind is also explicit in Miall *et al.* (1993, p. 210): 'Because the internal models must predict the expected outcome of movement in a form suitable to compare with the actual outcome, these internal models would operate in visual or egocentric coordinates.'

Recognizing that sensory cancellation via prediction does not *require* a comparison between similar kinds of states has several important consequences. First, as already noted, it means that we need not posit a separate comparator mechanism that calculates the difference between two similar states. Second, because it does not require that a quasi-perceptual (for vision, an 'imagistic') state be generated in order to accomplish sensory attenuation, it suggests a difference in where, neurologically, researchers should look when trying to locate the source of sensory attenuation disorders. In the case of vision, one would not necessarily investigate areas responsible for generating visual imagery, since, *pace* Grush, corollary discharge may not draw on imagery at all.

Of course, it remains an open empirical question whether the filtering hypothesis or the comparator hypothesis as I have described it is in fact correct. Both proposals are compatible with recent work that has explored the operation of efference-copy driven sensory cancellation in the visual system and its deficits in schizophrenic patients (Lindner *et al.*, 2005). They are also both compatible with Sperry and von Holst's classical work on the visual system.[15]

---

[14] 'Plant' is a control-theory term for the system being controlled, such as the body or the eye.
[15] Sperry (1950, p. 488) posits 'anticipatory adjustments', made only with regard to the direction and speed of each eye movement. von Holst (1950, p. 150) suggests thinking of efference copies as 'positive' anticipatory signals that, when meeting with an appropriate 'negative' incoming sensory signal at some early level of processing, cancel out the positive signal—

However, accepting the filtering picture in lieu of the comparator model has important consequences for our ability to extend such an account to explain inserted thoughts. According to the comparator model, where one's sense of agency is grounded in comparisons between similar kinds of states, for a *thought* to lack a sense of agency, there would need to be a prediction of that thought (i.e., an intention to think that *X*) that is compared with the actual thought (e.g., the thought that *X)* (Campbell (1999) develops a theory along these lines). A number of theorists have questioned whether we can plausibly suppose that there really are intentions to think *that X* corresponding to each actual thought *that X* (Gallagher, 2004). What would be the purpose of such a mechanism, given that one's cognitive system is not normally faced with the problem of distinguishing between other and self-generated thoughts?

Campbell hypothesizes that such mechanisms might allow one to continually monitor the semantic appropriateness of one's thoughts, allowing one to maintain a coherent train of thought. But, as Gallagher (2004, p. 12-13) argues, if the thoughts in question are already conscious, why isn't this enough to ensure that they are appropriately monitored? Indeed, why think that a sub-conscious 'thought-checking' mechanism would be in a better position to judge the semantic appropriateness of a thought than the conscious subject himself? Also, the envisioned role of these thought-predictions—keeping one's thought 'on track'—is quite different than verifying that the thoughts are self-generated, so it remains unclear how defects in this mechanism would lead to one's thoughts seeming to be generated by someone else. The filter model sidesteps these worries by removing the need to suppose we continually (albeit sub-consciously) predict what we are about to think. This is because the state accomplishing cancellation is not of the same basic kind of the one that is cancelled. In the next section I discuss how filtering, applied to the internal reafferences generated during inner speech, may work to imbue inner speech with a phenomenology of agency.

---

precisely as 4 cancels out -4 when the two are added—preventing the input from continuing into higher processing areas. Neither makes an explicit appeal to 'comparisons' taking place.

5. The Filtering Hypothesis and Inner Speech

The potential implications of the filter model for experiences of auditory hallucinations and inserted thoughts in schizophrenia can now be sketched: if inner speech is a kind of mental action involving internally generated 'reafferences' of its own—reafferences which may not be properly filtered—one might anticipate phenomenologies of the kind reported as inserted thoughts and 'soundless' voices. Just as normal perceivers can 'nudge' their eyeballs to achieve a certain 'fractured' phenomenology (forcing one to cite 'movement' in one's visual field, as opposed to the external world), so too may deficits in the efference copy mechanisms of schizophrenic patients lead to a fracturing of the phenomenology attending inner speech.

By a 'fracturing' of phenomenology, I mean the presence of a phenomenology normally associated with external changes in the environment that comes to seem as though it is only registering internal changes, due in part to its lack of coherence with other sensory and proprioceptive information. In the visual case, the wobbling visual field is a visual phenomenology normally associated with the entire environment moving (as in an earthquake) that comes to seem as though it is only registering internal changes, due its lack of coherence with the information available to other sense modalities (note: the relevant fracturing is in one's *overall* phenomenology, as it results from conflicting messages from different sense modalities).

Revealing theoretical and empirical evidence for how unfiltered inner speech could result in a similarly fractured phenomenology—and linking this fractured phenomenology to the phenomenology of thought agency—is the central concern of what follows.

5.1 Imagery, Covert Actions, and Internal Reafferences

The critical task is now to show how the abnormal inner speech of schizophrenic patients can be explained by analogy to the filtering model postulated for the

visual system. There is a *prima facie* difficulty with the analogy: inner speech does not involve the input and processing of new information from one's environment, and thus would appear to involve no reafferent inputs of the kind that must be filtered or attenuated. This problem has led some (Jeannerod and Pacherie, 2004; Proust, 2006; Grush, 2004) to search for ways to assimilate the mechanisms at play in perceptual and motor activity to those at work in 'covert' actions (such as action planning and imagination) as well.

While the notion of a 'reafference' is traditionally only applied to perceptual experience, Proust suggests that, from the standpoint of simulation theory, we have reason to posit something equivalent in the case of merely simulated or imagined actions. For Proust,

> There is no fundamental difference whether control applies to external or to internal actions. In both cases, the brain uses its own internal states and *stored reafferences* to simulate and regulate its own processes (2006, 4.1, emphasis added).

Jeannerod and Pacherie (2004, p. 125) work from a similar hypothesis, suggesting that 'actions, whether or not they come to execution, are centrally simulated by the neural network, and…this simulation is the basis for action recognition and attribution.' A key part of their view, as with Proust's, is that actions are simulated 'whether or not they come to execution'—thus, they are simulated in the absence of input from the environment.

According to these theorists (and Grush (2004)), an internally generated prediction state, expressed in visual or proprioceptive coordinates (i.e., a state of the same kind as sensory input states), is triggered both during the *mere* simulation of action *and* during normal perception (where there is reafferent input). I have already argued that, as an account of the mechanisms underlying perception, this sort of 'comparator' theory posits more than is needed to explain the observed attenuations of self-initiated sensory changes.[16] So, I disagree with

[16] It should be noted that the simulation account cannot explain how perceptual learning occurs any better than the 'filtering' approach I have suggested, since, on the simulation view, errors in goal-directed action are only correctable once there is a stored representation of what the correct

simulation theorists such as Proust and Jeannerod and Pacherie, who hold that 'internal' or 'stored' reafferences are triggered 'whether or not' planned action comes to fruition, since such states may not be needed for normal perception (unless these 'internal reafferences' are understood as being the kind of filter signals I described above).

Yet, when it comes to the class of thought we intuitively mark as 'imagery' or 'imagination' (where imagination essentially involves sensory imagery), the idea of quasi-perceptual 'internal reafferences' may well be relevant. In imaging, we are able to simulate acts of perception through the triggering of 'imagistic' representations—exploiting what, intuitively at least, seem like representations of the same basic kind as those activated in perception, and which carry the same kind of information about the external world (even if these kinds of representations are *not*, if I am right, necessarily deployed in normal perception as predictions that are compared with input). Indeed, a key difference between imagery and perception may be that only the former involves a willful generation of modality-specific, quasi-perceptual representations that serve as input into further cognitive processing.

The idea I shall now pursue is that inner speech is a kind of imagination or *mere* simulation of action (in this case, simulation of speaking), and, as such, can be seen as involving internally generated imagery in need of proper filtering or attenuation. But why think that imagery (and inner speech in particular) involves generating signals that serve as *input* to further processing—input that is normally attenuated or filtered? This important question can be broken into two: first, why think that imagery is in some way separate from normal thought in general, such that it could serve as input to later cognitive processing; second, why would this internally generated input normally be filtered or attenuated?

Fodor (1975, p. 175-195) provides the beginnings of an answer to the first question. Following Wittgenstein (1953, esp. *ff.* 137, fn. 2(b)), Fodor argues that images are too ambiguous to ground the intentional content we pre-reflectively

---

action *would have been*. The crucial learning question concerns how this 'correct' representation gets to be available to the cognitive system in the first place.

take them to have. That is, what we might intuitively think of as an image representing John standing could equally well represent that John is tall, or that John has two legs, or that John is skinny, and so on, for a variety of propositions.[17] Fodor concludes from this that, while images are realized in a different ('iconic') representational format than 'discursive' thought, they must nevertheless have a discursive description applied to them so that they can be properly disambiguated, and then appropriately used by a cognitive system. It remains an open question how representations in these two formats interact in the ways envisioned. But, for present purposes, it allows us to see why representations generated 'offline' during imagery tasks might themselves serve as inputs (in need of filtering, attenuation, or, in this case, interpretation) for later cognitive processing.

Of course, the representational format and informational richness of visual imagery is a matter of long dispute (see, e.g., Tye (1991), Pylyshyn (2002), Slotnick *et al.* (2005)); less commonly discussed are the perhaps more developed (and certainly more commonly exploited) resources of *auditory verbal imagery* involved in the inner speech of most humans. The imagery debate aside, it should be uncontroversial that verbal imagery is tied to sensation in roughly the same way as visual imagery, whatever the ultimate nature of this link may be. This link to sensation suggests that, if a modular theory of mind (Fodor, 1983) is generally correct, then inner speech involves interaction between central processing and an auditory or speech perception module. Here again is reason to think that this form of 'imagination' involves internally generated inputs (i.e. internal reafferences) that could, in normal cases, be filtered or attenuated.

Why think that this 'input' would normally be filtered? This is the second question above. An answer can be found in two places: research on the developmental roots of inner speech, and comparative neurological data during imagery and action/perception tasks, both of which suggest that (mere)

[17] The same indeterminacy is present if we take images to represent objects instead of states of affairs (i.e., to be the bearers of *reference*, but not truth). So long as images represent in virtue of resembling their referent, an image of John is equally an image of his twin, and so on. See Fodor's example involving 'a pinwheel sort of thing' (1975, p. 182-183).

imagination or simulation shares underlying mechanisms with perception and action. Following Jones and Fernyhough (2006)—themselves drawing on the empirical work of developmental psychologist Vygotsky (1934/1987)—we can see the very ability to engage in inner speech as grounded in the ability to simulate acts of auditory perception and speech 'offline'. On this view, we *first* learn to grasp linguistic meaning through hearing and watching others interact, then acquire the ability to speak. Only later are we able to engage in inner speech through simulating these learned abilities, exploiting a store of auditory verbal imagery in the process. If one's own normal speech is automatically accompanied by filter/efference-copy commands that attenuate one's sensitivity to the sound of one's own voice (as was found by Blakemore *et al.* (2003)), and if inner speech, due to its developmental roots, must draw on the same abilities and mechanisms as 'out loud' speech, then it is plausible suppose that inner speech is normally accompanied by filter/efference-copy signals as well.

Measurements of brain activity during perceptual and imagery tasks also bear out the hypothesis that a common set of mechanisms underlies both overt (e.g., speech, perception) and covert (e.g., inner speech, visual imagery) actions. Jeannerod and Pacherie (2004, p. 130) discuss results from neuro-imaging studies showing activity in a common neural network activated during both (mere) motor imagery and action preparation. Currie and Ravenscroft (1997) cite a wealth of similar data: for instance, by measuring regional cerebral blood flow, Roland and Friberg (1985) and Goldenberg *et al.* (1989) showed that common cortical regions were active both during visual perception and visual imagery; measuring alpha rhythms, Davidson and Schwartz (1977) showed that activity increased in the visual cortex during visual imagery. More recently, Slotnick *et al.* (2005) found similar visual imagery and perception tasks to evoke cortical activity with the same precise visual field topography ('retinotopy'), and to cause common effects in motion processing region MT+. Kosslyn and Thomson (2003) present a meta-analysis of a variety of neuroimaging studies with similar findings.

The other main source of neurological evidence that imagery and perception involve common mechanisms comes from patients with perceptual or

motor deficits that are mirrored in their abilities to use imagery. Sirigu *et al.* (1995) found that deficits in a patient's ability to move her fingers were attended by similar deficits in her ability to imagine moving her fingers. Corresponding data has been found regarding visual imagery and vision (Brooks, 1967): patients who, due to brain lesions, neglect a certain region of the visual field will neglect that same region in visual imagination. Another suggestive link stems from the well-known work of Kosslyn *et al.* (1978), who found that when subjects are asked to imagine making a movement, the time taken closely mirrors that taken in reality (and, similarly, that difficult motor-actions take longer to perform in imagination just as they do in reality), again implicating a common set of mechanisms. Notably, for schizophrenic patients experiencing inserted thoughts and voices, imagined movements tend to take the same time whatever the level of difficulty (Danckert *et al.*, 2002, p. 605-609).

Turning finally to the auditory imagery of inner speech implicated in auditory hallucinations and inserted thoughts, Stephane *et al.* (2001, p.75) present evidence that 'inner and external speech have the same developmental origin and both have perceptive and motor components.' To cite one striking example from their survey, Dell (1978) found that 'tongue-twisters', such as the phrase 'unique New York', cause the same sort of errors when 'uttered' only in inner speech as they do when said aloud. Looking at brain activity, Ford and Mathalon (2004) found a dampening of auditory cortex responsivity during both talking and inner speech in normal subjects, but found no such dampening to attend either activity in schizophrenic patients prone to auditory hallucinations.

The neurological and developmental evidence linking 'out-loud' speech to inner speech, combined with the view of inner speech as a kind of imagination involving the interaction between sensory modules and central processing, suggests the following picture. Episodes of inner speech are accompanied by efference copy signals just as normal speech is (due to the former drawing on the same underlying neural mechanisms and abilities as the latter); and to generate inner speech is to generate a kind of *imagery*, which requires interaction among multiple cognitive modules. For thought to become inner

speech, central processing must draw upon the resources an auditory memory module. It is the signal returning from this module that serves, in lieu of perceptual feedback, as the relevant 'reafferent' signal to be attenuated by the efference-copy triggered filter command. Defects in the filtering of this signal may lead to the abnormal phenomenology associated with thought insertion.

In closing this section, it should be noted that some have challenged the proposed link between inner speech and the 'voices' heard by schizophrenic patients. In a series of elegant experiments, Evans *et al.* (2000) showed schizophrenic patients prone to AVHs not to be impaired, relative to patients without AVHs, in various tasks involving the interaction of the 'inner voice' and 'inner ear'. In one experiment, patients prone to AVHs proved equally skilled to non-hallucinators at determining which pairs of written words were homophonic, which rhymed, and which 'sounded totally different' (they were asked to accomplish the task by silently 'pronouncing' the words in inner speech) (2000, p. 144). Evans *et al.* concluded that inner speech and auditory verbal hallucinations are not connected 'in a simplistic or direct way' (2000, p. 137). Yet, since it is well known that AVHs are significantly decreased by tasks that draw on or 'preoccupy' the normal mechanisms related to inner speech (Bick and Kinsbourne, 1987; Nelson *et al.*, 1991; Margo *et al.*, 1981), an alternative interpretation of their data is that the presented tasks sufficiently taxed the mechanisms underlying inner speech as to suppress normally defective corollary discharge responses. Evans *et al.* discuss this possibility themselves, noting that AVHs 'appear to be spontaneous and unbidden,' and, therefore, 'it is possible that our tasks did not tap into this potentially crucial feature of AVHs' (2000, p. 146). Thus, their findings do not close the door on inner speech-based theories of AVHs and inserted thoughts, even if they indicate that a mere appeal to 'deficient inner speech' is not explanatorily sufficient. In particular, an explanation of the intermittent character of the posited inner speech deficits will eventually be required. This issue is discussed further in section six below.

5.2 The Filtering Hypothesis and Inner Speech – Phenomenological Consequences of Filtering Deficits

So far I have explained how sensory attenuation may occur through efference-copy generated filter signals, and have given reason to think that the same principles can be extended to covert actions such as inner speech. It remains to be explained how a lack of filtering in the case of inner speech could result specifically in a phenomenology that leads one to think that one's own thoughts are someone else's. That is the project of this section.

A lack of filtering could give the phenomenology of inner speech an element of extraneity in a variety of ways. Just as with vision, there may be distinctive *structures* or *patterns of changes* that correlate with 'filtered' inner speech (and with hearing oneself speak) that are different from those associated with hearing others speak (the patterns being reversed in the case of inserted thoughts). Think again of the eye-nudging case: it is not that one's visual experience while poking one's eye is more or less vivid or rich than during normal visual exploration. The salient phenomenological difference is rather a structural difference in the nature of the respective signals across time—specifically, the difference consists in how similar successive states of the (downstream, conscious) visual system are to each other in the nudging and non-nudging scenarios. When efference copies are triggered during normal visual exploration, there will at times be considerable gaps (in terms of the type of information nested) between successive visual states, due to the filtering of the signal as the organism saccades (resulting in a 'stable' visual field). Conversely, the 'blurring' that occurs when efference copies are suppressed (in the eye-nudging case) results from the relative similarity of information nested in each successive state during active perception—put simply, less of the signal across time is 'thrown out'.

Similar structural differences may exist between filtered and unfiltered inner speech. One possibility is that filter signals normally prevent inner speech from being fully enunciated, both in order to speed its progression and to decrease the cognitive resources needed for its processing. When the filter signals are absent (due to an efference copy defect) one's inner speech may be

fully enunciated, slowing its progression and giving it an alien quality. By subtracting unnecessary information from the signal, filtering would enable normal inner speech to be easily suppressible in the event of unexpected perceptual input from outside (after all, even if central processing needs to draw on stored auditory information or abilities to get inner speech going, this does not mean that it needs everything that comes back from these modules). If filtering is lacking, inner speech may not be automatically 'overlooked' when other inputs arise, and may itself seem to be one of those externally-derived inputs, not a self-generated signal.[18]

These suggestions for what the precise phenomenological consequences of a lack of filtering of inner speech may be are only schematic. They serve principally as pointers toward the correct approach to understanding the phenomenology of agency. We face a major obstacle in explaining what it is like to have this phenomenology in any fine-grained way because most of us do not know what it is like to have such experiences. When it comes to sympathetically imagining inserted thoughts, we are in the same position that someone who has never rubbed his eyes (or experienced an illusion) would be in were he asked to imagine his visual field 'wobbling'. It seems it would be quite difficult for such a person to understand what was meant by a 'visual field' that can 'move' independent of objects moving in the environment.

This is overlooked in most discussions of the phenomenology of thought-agency; implicit in such discussions is the assumption that the *lack* of this phenomenology is something most people are familiar with or can easily imagine. The best we can do, I suggest, is point to structural and qualitative features this phenomenology may have, based on cases with which we are familiar. These features are still explanatory, however, since certain structural differences in phenomenology may plausibly be associated with the detection of changes in

[18] The 'hyperreflexivity' posited by Sass and Parnas (2003) and Gallagher (2004)—whereby the apparent extraneity of inserted thoughts arises out of schizophrenic patients becoming over-aware of their own thinking—may actually be a result of a failure of filter commands to meet and attenuate the signals returning from centers involved in generating auditory imagery. If that is right, then the putative hyperreflexivity consists not in an introspective faculty 'looking too hard,' but from the internally generated signal's unusual 'strength' (i.e., the increased amount of cognitive resources required for its processing).

one's environment, just as the visual field wobble is a visual phenomenology naturally associated with the visual perception of the entire room shifting to the side.

Notice that this account of the phenomenology of thought *agency*—and of its possible disorders—places that phenomenology within the auditory sense modality, as being embedded in certain filtered structures and patterns of auditory verbal imagery. Similarly, the phenomenology of agency can also at times be *visual* in nature, such as when intended eye movements result in efference copy commands that lead to a stabilized visual field; the filtering out of information received between the beginning and end of a saccade results in a kind of agency phenomenology that is correlated with willful acts of visual exploration. This kind of agency-phenomenology is not contained within imagistic 'snapshots', but rather arises out of distinctive patterns of information across time. If the input information between the beginning and end of a saccade is not filtered, the downstream pattern of information—across time—will be different; it will constitute a phenomenology associated with changes in one's environment.

Other theorists (Proust, 2006; Jones and Fernyhough, 2006) apply a similar efference-copy/reafference paradigm to explain how inner speech becomes mistaken as being externally generated, yet hold that an 'emotion of self-authorship' is the relevant aspect of phenomenology that is altered or missing due to efference copy defects. This is a crucial misstep, since it requires that we have some idea of what this emotion is in the normal case. But, as noted at the outset, it is far from clear what this emotion might be, and if it truly attends most normal thought. To suggest that it is normally unnoticed or taken for granted is simply to suggest that it is not part of non-pathological *experience*, and, hence, that it is not an element of phenomenology after all.

The appeal to an 'emotion of self-authorship' is motivated by comparator models that place the relevant 'mismatch' below the level of consciousness. On such accounts, it is only an unusual emotion (or lack thereof) that bubbles to the surface in schizophrenic experience, the patient's sensory phenomenology being

otherwise normal. The account I have given does not appeal to peculiar emotions, but rather to structures of changes in downstream sensory information—changes that are themselves present to consciousness as a part of what Gallagher (2004) calls 'first order phenomenology,' and which may plausibly lead to reports of thought insertion. This way, a *distinct* form of unusual phenomenology is described—one which sets schizophrenic cases apart from the normal lack of agency attending things such as melodies stuck in one's head. The phenomenology of agency that concerns us is therefore one that is embedded in all first order sensory and proprioceptive phenomenology as diachronic, action-sensitive patterns of information; it does not stand apart from them as an inscrutable emotion.

5.3. Fractured Phenomenologies – Unified Mechanisms for Inserted Thoughts and Auditory Verbal Hallucinations

Earlier I spoke of the way in which one's overall phenomenology becomes 'fractured' in the eye nudge case. The fracturing is due to a conflict between what the visual system is registering and what other sensory and proprioceptive inputs indicate. It is worth reflecting on the kind of fracturing that (if the present hypothesis is correct) takes place in the case of inserted thoughts. Interestingly, there does not seem to be a *direct* auditory equivalent of a visual field; that is, there is no 'auditory field' that seems to wobble when one has an auditory experience as of someone speaking, while the inputs of other sensory channels indicate there is no such speech occurring. We can suppose, however, that inner speech and the perception of one's own out-loud speech are filtered in various (attenuating) ways that set them apart, phenomenologically, from the perception of the speech of others (this may be part of the reason our voices sound different to us when played back from a recording). When this filtering is lacking, inner speech may be characterized by the kind of unfiltered phenomenology normally attending the perception of another's speech.

Neurological data backs up these speculations. Ford and Mathalon (2004) found that, in schizophrenic patients prone to AVHs, responsivity of auditory

cortex during inner speech is similar to that during normal auditory perception (in normals inner speech is accompanied by a dampening of responsivity in auditory cortex). Dierks *et al.* (1999) found that when auditory hallucinations occur (as indicated by the patient) brain metabolism increases in the primary auditory cortex on the left side, indicating that the nervous system in these patients behaves as it would during normal speech perception (they also note that inner speech is normally attended by a decrease in primary auditory cortex metabolism). And Daprati *et al.* (1997, p. 73) found that 'during verbal hallucinations in schizophrenic patients…the sensory areas for language remain active, which suggests that the [normal] cancellation process does not operate.' Assuming that, when this lack of cancellation occurs, other sensory and proprioceptive inputs conflict with the notion that someone nearby is speaking, a fracturing of one's overall phenomenology results.

Yet there is also a *second* fracturing that can occur in the phenomenology of inner speech. The second fracturing has to do with the comparative lack of richness in sensory character attending unfiltered inner speech. Because the sensory character of inner speech is not nearly as rich as that of actual speech perception, an act of inner speech is usually phenomenologically distinguishable *as inner speech* partly due to this paucity of sensory character. Yet, when this lack of sensory character is coupled with a lack of normal filtering, the inner speech at once possesses an element of innerness (the lack of sensory character) and outerness (the lack of filtering). This constitutes a second form of fracturing.

These two forms of fracturing would place unfiltered inner speech in a very peculiar phenomenological category. Experiencing such inner speech would be somewhat like hearing another speak, and somewhat like engaging in inner speech, but not exactly like either. How would one describe such an experience? Perhaps as an 'inserted thought' (due to its lack of sensory character), yet it seems equally likely that one might describe it as an episode of seeming to hear someone speak (due to its lack of filtering). My suggestion, therefore, is that AVHs and inserted thoughts may at bottom be the same phenomenon reported

in different ways, precisely because their phenomenology straddles the fence between speech perception and inner speech.

The usual approach to understanding these phenomena is to assume we have a reasonable idea of what AVHs are like (they are just like hearing someone speak) which leads to a puzzle over what might account for reports of inserted thoughts. I suggest instead that we question whether most AVHs really are just like hearing someone else speak, and instead assimilate them to the fractured phenomenological realm just described. Given the co-presence of these putatively distinct symptoms in many patients (Nayani and David, 1996), a unified account of their source is desirable. Crucially, this is a unification that takes seriously the reports of patients who claim their inserted thoughts are *not* simply like cases of hearing someone speak.

6. Loose ends: Intermittency and Content

There are two important characteristics of inserted thoughts that have not been explained by the present account. The first is their intermittency. Presumably, not *all* of the inner speech of schizophrenic patients is improperly filtered and lacking in the usual phenomenology of agency. But if the inserted thoughts are caused by a deficit in filtering, why wouldn't this deficit afflict all of patients' inner speech? The short answer is that these filtering errors must themselves be intermittent. Admittedly, this leaves the intermittent character of inserted thoughts unexplained. But this should not worry us, as it is a normal feature of explanations of intermittent phenomena that they first locate the causal factors, and then *separately* give an account of why these causal factors are only intermittently active. For instance, my intermittently runny nose may be explained by the presence of high levels of histamines in my bloodstream. Once we know that histamines are responsible for my runny nose, we can then ask why these histamines are only occasionally present in the high degree needed to induce a runny nose. A separate account is then given explaining how contact

with intermittently present allergens leads to the release of antibodies which in turn cause an increase in histamine levels.

Note that a key part of arriving at the explanation of intermittency in this case is that the biological basis of histamines is properly understood, for this is what allows the histamine explanation to be linked to a further explanation involving antibodies and allergens. Similarly, we should expect that, once the biological or neurological basis of the posited filtering mechanisms is located, we will be in a much better position to explain why these mechanisms are only intermittently unreliable.

A second worry may be that the present account does nothing to explain the particular content of inserted thoughts—i.e., why do such thoughts have one subject-matter and not another? If it were the case that all or most inserted thoughts had a particular content—were personally abusive, say—there would be good reason to suppose that this content plays a causal role in generating the relevant abnormal phenomenology. As reports of the contents of large numbers of actual inserted thoughts (as opposed to auditory verbal hallucinations) are hard to come by, it is difficult to assess this hypothesis. We do know, however, that AVHs often have a positive or comforting content for their bearers (Bleuler 1911/1950). When asked whether they ever experienced 'pleasant' voices, forty-eight of 100 schizophrenic patients surveyed by Nayani and David indicated that they did (1996, p. 182). Nayani and David also found that, while the development of coping strategies by patients was associated with the presence of pleasant voices, this change in content was not sufficient to stop the hallucinations (1996, p. 186). Thus, any account which tries to explain inserted thoughts in terms of anxiety-ridden or displeasing content will have to give a separate account of the nature of AVHs. A virtue of the filtering account is that it can treat inserted thoughts and AVHs as a single experiential phenomenon. The filtering hypothesis's approach to the question of content will be to predict no special link between thought content and a lack of a sense of agency.

By avoiding an appeal to content as a causal factor, the filtering approach remains broadly applicable in a way that allows it to explain other schizophrenic

symptoms as well, such as delusions of control (where the subject feels her bodily movements not to be under her own control), and the well-documented failures to attenuate the consequences of self-initiated changes in a variety of sense modalities (Shergill *et al.*, 2005). An account of inserted thoughts that focuses on their content as causally relevant (e.g., Vosgerau and Newen (2007, p. 38-40)) could not extend itself to these cases in any clear way. Given that inserted thoughts rarely occur in isolation, but rather are just one of several typically co-present symptoms of schizophrenia, it is highly desirable to develop an explanation applicable to more than one of these co-present symptoms.

7. Virtual Reality, 'Presence', and Belief Fixation

With this proposal in hand, how close are we to a solution to the puzzle of extraneity? One might worry that, despite having a firmer grip on the possible phenomenology and etiology of thought insertion and auditory hallucinations, it is still hard to see what would cause schizophrenic patients to continually form such delusional *beliefs*. If their general ability to reason is not impaired, why do patients so strongly resist concluding that they are simply perceptually unreliable, persisting instead in attributing the agency of their thoughts and 'voices' to someone else? After all, it seems one can imagine having experiences like those of schizophrenic patients without thereby coming to believe that someone else is speaking or inserting thoughts into one's mind. Would we not, in such circumstances, simply conclude that we'd become perceptually unreliable?

Part of what accounts for such puzzlement may be an unrealistic view of the relation between perceptual phenomenology and belief. When we imagine it possible to have the same perceptual phenomenology as schizophrenic patients while lacking their beliefs, we tacitly assume that perceptual phenomenology is cleanly separable from belief. However, the fact that sensory modules process information prior to its involvement in belief fixation does not entail that *perceptual phenomenology* (a conscious, and relatively late state in the information-processing chain) can itself completely float free of belief. It is not as

though we normally *choose* what to believe based on perceptual phenomenology; perceptual phenomenology just is a kind of belief—albeit, one that can be trumped in special circumstances by extenuating background beliefs.

If perceptually-formed beliefs are in large measure formed pre-consciously, how does belief relate to early perceptual input? This is a central question of researchers developing virtual environments (VEs). Since present technologies involved in the creation of VEs cannot approach the level of informational detail and computational speed needed to create a full-on replication of the sensory consequences of perceiving a genuine environment, a crucial task is to discover the minimal perceptual cues needed in order to make a VE convincingly 'present' to a perceiver. 'Presence' is a technical term in the VE literature used to refer to a VE's ability to instill in the perceiver a sense of actually 'being there' in the VE. Presence is measured in a variety of ways, the two most common being questionnaires (given to subjects after the fact), and measurements of physiological markers such as EKG and galvanic skin response during the exploration of a VE (Sanchez-Vives and Slater, 2005).

Most VEs require subjects to wear a head-mounted display (HMD) which funnels computer-generated visual information to the left and right eyes via small screens (each eye receives slightly different signals in order to induce the proper three-dimensional effect). Crucial for presence in a VE is that the rate at which graphics are refreshed is at least 15Hz, and that these changes occur in accordance with the subject's movements (Sanchez-Vives and Slater, 2005; Barfield and Hendrix, 1995). For instance, in a task where subjects were asked to retrieve an object from a (virtual) room with a deep 'pit' at its middle (leaving only a foot or so of clearance on any side of it for one to make one's way around to the goal object), heart rates were significantly increased when the frame-refresh-rate was over 15Hz, yet did not significantly change when frame rates were lower (here heart rate is being used to judge presence) (Meehan *et al.*, 2002).

Interestingly, the visual realism of a virtual environment has been shown to be a negligible factor for the creation of presence (Usoh *et al.,* 2000; Zimmons

and Panter, 2003; Sanchez-Vives and Slater, 2005). Crudely rendered objects, persons, and rooms can achieve high ratings on all tests of presence, so long as the screen shots of such environments are updated at the required rate, and are done so pursuant to the actual head and body movements of the subject. Two other factors shown to significantly increase presence are a representation of the body of the subject herself (even if crudely or oddly rendered), and the ability of a subject to actually walk about (as opposed to only looking) while engaging with the VE (Sanchez-Vives and Slater, 2005).

These findings have at least two important consequences for thinking about the puzzle of extraneity. The first is that our pre-theoretical intuitions about the factors in perception that lead to the fixation of belief may not be reliable—a richly detailed environment may be relatively unimportant for the formation of belief-like states, so long as certain patterns of information are received at the proper rate by the early perceptual system. Second, this work suggests a strong link between the sense of *extraneity* one has regarding a VE and the receiving of certain action-sensitive patterns of sensory stimulation—patterns that may arise partly out of specific filtering algorithms being applied (or not). Even though the subjects 'know better' than to become nervous in front of an unhappy-looking audience[19], or to fear the virtual 'pit' in the center of a room, a degree of belief that these scenarios are real nevertheless attaches, resulting in increased heart rates and impeded performance. As the technologies underlying VEs become more sophisticated, opportunities will arise to further test the filtering hypothesis, as perceptual feedback from the VE could conceivably be manipulated in ways to duplicate in normal subjects the effects postulated to occur in patients with efference copy deficits.

[19] Sanchez-Vives and Slater (2005) asked their subjects to deliver a speech to two different virtual audiences. One was 'positive'—the listeners' eyes programmed to remain open and fixed on the subject's, the virtual characters making nodding movements and sounds of approval throughout the talk. The other was 'negative'—audience members avoided eye contact and assumed postural positions indicating a lack of interest. The virtual humans in the 'audience' were very primitively rendered, showing only basic facial expressions and assuming relatively stiff, mechanical postures. Yet, even experienced speakers among the subjects reported being flummoxed by the 'negative' audience, finding themselves emotionally upset and performing far worse than normal, while speakers addressing the 'positive' audience performed well and found their experiences gratifying.

More importantly, VEs may eventually offer a new form of therapy for those suffering from the hallucinations and delusions characteristic of schizophrenia. VEs have already been used to help treat some phobias. Patients are submerged in virtual environments simulating their phobias to various degrees of intensity, allowing them to gradually become accustomed to facing such situations without anxiety (Rothbaum *et al.*, 1999). Specific perceptual-motor deficits—such as those attending Parkinson's Disease—are also treated using VEs (Prothero, 1993). While the VE does not stimulate increased production of dopamine, it does allow for the discovery of new coping strategies by presenting disability-specific challenges (Wann, 1996).

These two kinds of therapy could be combined in the treatment of schizophrenic symptoms. The sort of 'unfiltered' sensory feedback thought to attend some actions of schizophrenic patients can be independently generated by manipulating the behavior of the VE as it is explored, allowing the patient to become more familiar with certain types of dissociations. Becoming accustomed to irregular feedback in a VE may both decrease the patient's anxiety in the face of similar pathological experiences, and may allow for the discovery of novel coping strategies—strategies that can only be discovered in an atmosphere lacking the anxiety typically surrounding pathological experience. Even if a VE cannot duplicate the phenomenology of unfiltered inner speech (since the only inputs are normal sense organs), growing accustomed to unfiltered visual and auditory input (in situations where efference copies should lead to filtering) may still allow for the development of coping strategies relevant to dealing with unfiltered 'internal' stimuli.

8. Conclusion

There is no short or obvious path toward answering the puzzle of extraneity. The overall strategy adopted herein has been to take the first person reports of schizophrenic patients seriously, and to search for the simplest way of understanding them in terms of the mechanisms underlying normal perception

and thought. I have given reasons to think that thought insertion and auditory hallucinations may result from single general deficiency—a lack of the proper 'filtering' of inner speech—and, therefore, that these two cardinal symptoms of schizophrenia may themselves best be thought of as marking distinctions in degree of a single underlying cognitive disorder. Importantly, the mode of explanation employed—invoking efference copy-driven sensory cancellation—is one that can be extended to explain schizophrenic hallucinations and delusions of control within other proprioceptive and sense modalities as well.

On the view I have proposed, we should locate the phenomenology of agency *within* the traditional sensory and proprioceptive modalities, as diachronic patterns of high-level processing, not outside of them as an extraneous emotion or 'feeling of effort'. Such patterns *just are* the extraneity that attends schizophrenic experience and which, in non-pathological cases, allow one to distinguish externally from internally caused changes in sensation. Understanding why these patterns result in the specific delusions and hallucinations of schizophrenia, I have argued, requires looking at the way overall phenomenologies are shaped by the interaction among multi-modal sensory and proprioceptive inputs. Doing so allows one to see the fault lines along which phenomenologies may fracture.

Department of Philosophy
The Graduate Center of the City University of New York

References

Barfield, W. and Hendrix, C. 1995: The effect of update rate on the sense of presence within virtual environments. *Virtual Reality: The Journal of the Virtual Reality Society*, 1, 3-16.

Bick, P.A. and Kinsbourne, M. 1987: Auditory hallucinations and subvocal speech in schizophrenic patients. *American Journal of Psychiatry*, 144, 222-225

Blakemore, S.J., Wolpert, D.M., and Frith, C.D. 1998: Central cancellation of self-produced tickle sensation. *Nature Neurosciences*, 1, 635-640.

Blakemore, S.J., Wolpert, D. and Frith, C.D. 2002: Abnormalities in the awareness of action. *TRENDS in Cognitive Science,* 6, 237-241.

Blakemore, S.J., Oakley, D.A. and Frith, C.D. 2003: Delusions of alien control in the normal brain. *Neuropsychologia,* 41, 1058-1067.

Blakemore, S.J. 2005: Recognizing the sensory consequences of one's own actions and delusions of control. In T. Feinberg and J. Keenan (eds.), *The Lost Self*. Oxford: Oxford University Press.

Bleuler, E. 1911/1950: *Dementia Praecox or the Group of Schizophrenias,* trans. J. Zinkin. New York: International Universities Press.

Brindley, G. and Merton, P.A. 1960: The absence of position sense in the human eye. *Journal of Physiology (Lond.),* 153, 127-130.

Brooks, L. 1967: The suppression of visualization by reading. *Quarterly Journal of Experimental Psychology*, 19, 289-299.

Campbell, J. 1999: Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist,* 82, 609-625.

Campbell, J. 2002: The ownership of thoughts. *Philosophy, Psychiatry, and Psychology*, 9, 35-39.

Campbell, T. and Pettigrew, J. 2004: Testable corollaries, a conceptual error, and neural correlates of Grush's synthesis. *Behavioral and Brain Sciences*, 27, 398-400.

Chapman, L.J. and Chapman, J.P. 1988: The genesis of delusions. In T. Oltermanns and B. Maher (eds.), *Delusional Beliefs*. New York: Wiley.

Currie, G. and Ravenscroft, I. 1997: Mental simulation and motor imagery. *Philosophy of Science*, 64, 161-180.

Danckert, J., Rossetti, Y., d'Amato, T., Dalery, J., and Saoud, M. 2002: Exploring imagined movements in patients with schizophrenia. *Neuroreport*, 13, 605-609.

Daprati, E., Franck, N., Georgieff, N., Proust, J., Pacherie, E., Dalery, and Jeannerod, M. 1997: Looking for the agent: an investigation into consciousness of action and self-consciousness in schizophrenic patients. *Cognition*, 65, 71-86.

Davidson, R. and Schwartz, G. 1977: Brain mechanisms subserving self-generated imagery: electrophysiological specificity and patterning. *Psychophysiology*, 14, 598-602.

Dell, G.S. 1978: Slips of the mind. In M. Paradis (ed.), *The first LACUS Forum*.

Columbia, SC: Hornebeam Press.
Dierks, T., Linden, D.E.J., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., and Singer, W. 1999: Activation of Heschl's gyrus during auditory hallucinations. *Neuron,* 22, 615-622.
Evans, C.L., McGuire, P.K, and David, A.S. 2000: Is auditory imagery defective in patients with auditory hallucinations? *Psychological Medicine,* 30, 137-148.
Fodor, J.A. 1975: *The Language of Thought*. New York: Thomas Crowell.
Fodor, J.A. 1983: *Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.
Ford, J.M. and Mathalon, D.H. 2004: Electrophysiological evidence of corollary discharge dysfunction in schizophrenia during talking and thinking. *Journal of Psychiatric Research,* 38, 37-46.
Ford, J.M. and Mathalon, D.H. 2005: Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *International Journal of Psychophysiology*, 58, 179-189.
Frith, C.D. 1992: *The Cognitive Neuropsychology of Schizophrenia*. Hove, UK: Lawrence Erlbaum.
Frith, C.D. 2005: The neural basis of hallucinations and delusions. *Comptes Rendus Biologies,* 328, 169-175.
Frith, C.D. and Done, D.J. 1989: Experiences of alien control in schizophrenia reflect a disorder in the central monitoring of action. *Psychological Medicine,* 19, 359-63.
Gallagher, S. 2000: Self-reference and schizophrenia: a cognitive model of immunity to error through misidentification,' In D. Zahavi (ed.), *Exploring the Self: Philosophical and Psychopathological Perspectives on Self-experience.* Amsterdam and Philadelphia: John Benjamins.
Gallagher, S. 2004: Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology*, 37, 8-19.
Goldenberg, G., Podreka, I., Uhl, F., Steiner, M., Willmes, K., and Deecke L. 1989: Cerebral correlates of imagining colours, faces, and a map. *Neuropsychologia*, 27, 641-664.
Gould, L.N. 1949: Auditory hallucinations and subvocal speech: objective study in the case of schizophrenia. *Journal of Nervous and Mental Disease,* 109, 418-427.
Graham, G. and Stephens, G.L. 2000: *When Self-Consciousness Breaks.* Cambridge, MA: MIT Press.
Green, P. and Preston, M. 1981: Reinforcement of vocal correlates of auditory hallucinations by auditory feedback: case study. *British Journal of Psychiatry,* 139, 204-208.
Grush, R. 2004: The emulation theory of representation: motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377-442.
Haggard, P. 2003: Conscious awareness of intention and of action. In J. Roessler, J. and N. Eilan, (eds.), *Agency and Self-awareness*. Oxford: Oxford University Press.
Helmholtz, H. 1866: *Handbuch der Physiologischen Optik.* Leipzig: Voss.
Jeannerod, M. and Pacherie, E. 2004: Agency, simulation and self-identification.

Mind and Language, 19, 113-146.
Jones, S.R. and Fernyhough, C. 2006: Thought as action: inner speech, self-monitoring, and auditory verbal hallucinations. *Consciousness and Cognition* (forthcoming).
Kosslyn, S.M.1978: Measuring the visual angle of the mind's eye. *Cognitive Psychology*, 10, 356-389.
Kosslyn, S.M., Thompson, W.L. 2003: When is early visual cortex activated during visual mental imagery. *Psychological Bulletin*, 129, 723-746.
Lindner, A., Their, P., Kirchr, T.T.J., Haarmeier, T., Leube, D.T. 2005: Disorders of agency in schizophrenia correlate with an inability to compensate for the sensory consequences of actions. *Current Biology,* 15, 1119-1124.
Maher, B.A. 1974: Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30, 98-113.
Margo, A., Hemsley, D.R., and Slade, P.D. 1981: The effects of varying auditory input on schizophrenic hallucinations. *British Journal of Psychiatry,* 139, 122-127.
Meehan, M. Insko, B., Whitton, M. and Brooks, F.P. 2002: Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics*, 21, 645-652.
Mel, B.W. 1988: MURPHY: A robot that learns by doing. In D.Z. Anderson (ed.), *Neural information processing systems,* American Institute of Physics.
Mellor, C.H. 1970: First rank symptoms of schizophrenia. *British Journal of Psychology*, 117, 15-23.
Miall, R.C., Weir, D.J., Wolpert, D.M., and Stein, R.C. 1993: Is the cerebellum a smith predictor? *Journal of Motor Behavior*, 25, 203-216.
Nayani, T.H., David, A.S. 1996: The auditory hallucination: a phenomenological survey. *Psychological Medicine,* 26, 177-189.
Nelson, H.E., Thrasher, S., and Barnes, T.R.E. 1991: Practical ways of alleviating auditory hallucinations. *British Medical Journal,* 302, 327.
Prothero, J.D. 1993: The treatment of akinesia using virtual images. *Master's thesis*, Industrial Engineering, University of Washington.
Proust, J. 2003: Thinking of oneself as the same *Consciousness and Cognition*, 12, 495-509.
Proust, J. 2006: Agency in schizophrenia from a control theory viewpoint. In N. Sebanz and W. Prinz (eds.) *Disorders of Volition*. Cambridge, MA: MIT Press.
Pylyshyn, Z. 2002: Mental Imagery: In search of a theory. *Behavioral and Brain Sciences*, 25, 157-182 (-237 including commentaries and reply).
Roland, P. and Friberg, L. 1985: Localization of cortical areas activated by thinking. *Journal of Neurophysiology*, 53, 1219-1243.
Rothbaum, B., Hodges, L, Alarcon, R., Ready, D., Shahar, F., Graap, K., Pair, J., Herbert, P., Gotz, D., Wills, B., and Baltzell, D. 1999: Virtual reality exposure therapy for PTSD Vietnam veterans: a case study. *Journal of Traumatic Stress*, 12, 263–71.
Sanchez-Vivesm M. and Slater, M. 2005: From presence towards consciousness. *Nature Reviews Neuroscience,* 6, 332.

Sass, L.A. and Parnas, J. 2003: Schizophrenia, consciousness, and the self. *Schizophrenia Bulletin*, 29, 427-444.
Schneider, K. 1959: *Clinical Psychopathology*. New York: Grune and Stratton.
Shergill, S.S., Samson, G., Bays, P.M., Frith, C.D., Wolpert, D.M. 2005: Evidence for sensory prediction deficits in schizophrenia. *American Journal of Psychiatry,* 162, 2384-2386.
Sirigu, A., Cohen, L., Duhamel, J.R., Pillon, B., Dubois, B., Agid, Y., Pierrot-Deseilligny, C. 1995: Congruent unilateral impairments for real and imagined hand movements. *NeuroReport*, 6, 997-1001.
Slotnick, S., Thompson, W., and Kosslyn, S. 2005: Visual mental imagery induces retinotopically organized activation of early visual areas. *Cerebral Cortex,* 15, 1570-1583.
Spence, S. 1996: Free will in the light of neuroplasticity. *Philosophy, Psychiatry, and Psychology,* 3, 75-90.
Sperry, R.W. 1950: Neural basis of the spontaneous optokinetic response produced by visual inversion. *The Journal of Comparative and Physiological Psychology*, 43, 482-489.
Stephane, M., Barton, S. and Boutros, S. 2001: Auditory verbal hallucinations and dysfunction of the neural substrates of speech. *Schizophrenia Research*, 50, 61-78.
Strawson, G. 2003: Mental ballistics *or* the involuntariness of spontenaity. *Proceedings of the Aristotelian Society*, 103, 227-56.
Tye, M. 1991: *The Imagery Debate*. Cambridge, MA: MIT Press.
Usoh, M., Catena, E., Arman, S. and Slater, M. 2000: Using presence questionnaires in reality. *Presence-Teleoperators and Virtual Environments*, 9, 497-503.
Vosgerau, G. and Newen, A. 2007: Thoughts, motor actions, and the self. *Mind and Language*, 22, 22-43.
von Holst and Mittelstadt, H. 1950/1973: Das Reafferenzprinzip. *Naturwissenscaften,* 37, 464-76.
English translation, 1973: The reafference principle. In R. Martin (trans), *The Behavioral Physiology of Animals and Man. The Collected Papers of Erich von Holst,* trans. R. Martin. Miami, FL: University of Miami Press.
Vygotsky, L.S. 1934/1987: *Thought and language.* Cambridge MA: MIT Press. Original work published in 1934.
Wann, J. 1996: Virtual reality environments for rehabilitation of perceptual-motor disorders following stroke. *Proceedings of the First European Conference on Disability, Virtual Reality and Associated Technologies,* 1, 233-238.
Weiskrantz, L., Elliot, J., and Darlington, C. 1971: Preliminary observations of tickling oneself. *Nature*, 230, 598-99.
Wittgenstein, L. 1953: *Philosophical Investigations*. 2nd ed. trans. G.E.M. Anscombe. Oxford: Blackwell.
Wolpert, D. and Ghahramani, Z. 2000: Computational principles of movement neuroscience. *Nature Neuroscience Supplement*, 3, 1212-1217.
Wolpert, D., Miall, R.C., and Kawato, M. 1998: Internal models in the

cerebellum. *TRENDS in Cognitive Science*, 2, 338-347.
Zaretsky, M. and Rowell, C.H.F. 1979: Saccadic suppression by corollary discharge in the locust. *Nature,* 280, 583-85.
Zimmons, P. and Panter, A. 2003: The influence of rendering quality on Presence and task performance in a virtual environment. In *Proceedings of the IEEE Virtual Reality Conference 2003.* IEEE Computer Society, 293.