The comparator account on thought insertion, alien voices and inner speech: some open questions

Abstract

Recently, many philosophers and psychologists have claimed that the explanation that grounds both passivity phenomena in the cognitive domain and passivity phenomena that occur with respect to overt actions is, along broad lines, the same. Furthermore, they claim that the best account we have of such phenomena in both scenarios is the "comparator" account. However, there are reasons to doubt whether the comparator model can be exported from the realm of overt actions to the cognitive domain in general. There is a lingering worry concerning such explanations of thought insertion: the "What is compared to what?" problem. Here I examine two ways to tackle this problem. First: thought insertion consists of the misattribution of strings of inner speech which are not attenuated (thought insertion is thus another name for auditory verbal hallucinations). Second: thought insertion is misattributed inner speech which exhibits the same phenomenological characteristics as normal inner speech. After explaining the types of problem that each of these potential solutions faces, I conclude with a set of open questions that the comparator theorist has to tackle.

Introduction

Many philosophers and psychologists have recently claimed that the explanation that grounds passivity phenomena in the cognitive domain and passivity phenomena that occur with respect to overt actions is, along broad lines, the same. That is, what goes wrong in both cases is the same kind of thing. Many of those authors also hold that the explanation for the sense of agency allegedly common to some thinking (such as judging) and to overt actions is, at some interesting level, identical. This sort of position is developed in the writing of philosophers such as Campbell (1999), Peacocke (2007) and Proust (2006), and of psychologists such as Feinberg (1978), Frith (1992) and Ford et al. (2007), to name but a few.

In the case of the psychologists, the assimilation of agentivity in thinking to that in acting is motivated in good part by the search for a unified cognitive theory of schizophrenia. Among the many symptoms that can affect schizophrenia patients, there are outstanding ones such as delusions of control, thought insertion and auditory verbal hallucinations (AVHs). The psychologists that pursue this line strive to show that these three phenomena—and others, such as anarchic hand syndrome—which are all lived by subjects as experiences of passivity, can be explained as breakdowns in the same kind of mechanism. The philosophers in turn strive for a unified explanation of the sense of agency, since, given that we are said to have some feeling of agency in both realms—the cognitive and the physical—it is desirable to establish that there is only one sense of agency in play. Such a position would be strengthened if it turned out that the sense of agency is underpinned by the same kind of mechanism in both cases.

It seems that the most promising account of why we feel our actions and our thinking as our own, and of why some subjects feel either their actions or their thinking as alien, is

some version of the "comparator" account, forcefully put forward by Frith (see, e.g., Frith, 1992). Frith's theory draws on the corollary discharge model of perception proposed by Helmholtz (1860), and extended by von Holst & Mittelstaedt (1950) and Sperry (1950) to deal with motor acts. The model works in the following way: whenever a motor command is issued, the brain predicts, based on an efference copy (or corollary discharge) and the work of some "forward models", what proprioceptive and sensory feedback will ensue. If the prediction matches the actual feedback received, that feedback is attenuated and perceived as generated by the subject: the subject gets the feeling that the action is his own. If the incoming feedback signal that results from the motor act was not adequately predicted, it is felt in its full intensity and the subject does not get the sense of being the agent responsible for the movement. Frith's idea is that what goes wrong in passivity phenomena in general is that the forward models which issue predictions malfunction.

There are now several versions of the comparator account, as well as an increasing suspicion that it requires some sophistication or to be complemented by other accounts (see, e.g. Proust, 2006, Pacherie, 2007, 2011, Synofzik and Vosgerau, 2008, Carruthers, 2012, Frith, 2012). Here I focus on its applicability to the cognitive domain. As I say, the comparator account is supposed to apply to all kinds of passivity phenomena (and in particular, the passivity experienced by schizophrenia patients), and it is supposed to explain the sense of agency experienced both in the mental domain—when judging, trying, reasoning, imagining, etc. (see, e.g. Peacocke, 2007)—and in that of motor acts. However, there are reasons to doubt whether the comparator model applies to the cognitive domain. There is a lingering problem with explaining thought insertion, which

has led even Frith (2012)—but not many other authors—to hold that thought insertion cannot be accounted for by his theory. Here I explore and evaluate two possible solutions to that problem.

I will proceed as follows: first, I present the comparator model and explain the reason why it is, in principle, difficult to apply it to thought insertion—this is the reason Frith and others have in mind—specifically, that we do not know what is compared to what when we move from bodily actions to thinking. Then I examine two possible ways of meeting the challenge thought insertion represents. One of them consists of equating thought insertion to AVHs (see Langlad-Hassan, 2008). The other consists of explaining thought insertion as a misattribution of inner speech (I explain the difference between these two positions below). I hold that both approaches are problematic and that the successful development of either of them requires comparator theorists to tackle some questions to which they seem to lack straightforward answers. I also maintain that it is necessary for the comparator approach to revisit views on attenuation and the role of this notion within the theory. Which of the two options above (thought insertion equals AVHs/ thought insertion is misattributed inner speech) we should opt for depends on the view we adopt regarding attenuation in the cognitive domain. If we think that it makes sense that endogenously produced and consumed signals are attenuated as a result of successful comparisons, then we should hold that thought insertion is a case of AVHs by another name (which, we will see, is a problematic move in itself). However, if, for instance, we think that private signals are already attenuated because they are simulated signals, then we had better believe that thought insertion is *only* misattributed inner speech and dispute the current most common treatment of AVHs based on the comparator model—which

means that it may well turn out that the deeply problematic phenomenon for the comparator model is not thought insertion, but AVHs.

The comparator model

The comparator model draws on the idea that normal perception involves a corollary discharge or efference copy that is used to compute predictions about the sensory feedback that will be received after movement. The account generalizes the activation of this corollary discharge to all kinds of movement. Thus, any intentional movement of our body involves the following steps: first, we form the intention to move our body; then, we issue a motor command, a copy of which—the efference copy—is fed to the forward models; the forward models make a prediction, before the movement takes place, about the feedback that is expected as a result of moving; and finally, the prediction is compared to the actual feedback received from the movement. As occurs in perception, if the actual feedback matches the prediction, then that feedback is attenuated; otherwise, the full intensity of the feedback is experienced. The attenuation of signals is important, because it helps to distinguish changes produced by the self from changes produced by an external source. One important source of evidence in favor of this model is the phenomenon of self-tickling in schizophrenic subjects (see Blakemore et al., 2000); most people cannot tickle themselves, whereas some schizophrenic subjects can. Self-ticking can be explained as a lack of attenuation of the incoming sensory signals, which in turn can be explained as some failure in the forward models responsible for forming predictions.

The comparator model has another dimension, which—although it will not concern us here—it is interesting to mention, as it lends more plausibility to the whole approach. The model requires that predictions are compared not only to outcomes, but also to desired states. That is, we keep in mind the state we want to achieve and compare it with the predicted outcome. If there is some discrepancy between the two, then we correct our movements. Failures in our predictions thus lead to movements that do not achieve our goals—because the movements are not corrected. In a nutshell, the comparator model provides the resources to explain not only the phenomenon of attenuation, but also the mechanics of correction and control.

Frith and colleagues place much importance on the phenomenon of attenuation[1]. I have mentioned that attenuation is important because it is said to help the subject discriminate movements which are endogenously produced from exogenously produced movements. This leads Frith and colleagues to relate attenuation to the sense of agency. However, it is not clear how they think they are related. I take it that most comparator theorists hold that attenuation is necessary for the feeling of agency. At any rate, this is what one can glean from quotes such as the following: "Over-activity of the inferior parietal cortex might reflect a heightened response to the sensory consequences of movements the schizophrenia patients made during the scan, *contributing* to the feeling that movements are externally controlled" (Blakemore and Frith, 2003; 649, my italics); and "This could cause a lack of attenuation of the sensory consequences of self-produced actions, which

[1] A word about 'attenuation': in some writings, and in some authors, attenuation is used in the sense of "cancellation", such that everything that is successfully predicted about the incoming signal is cancelled, i.e. not felt. However, I will understand that an attenuated signal is not a cancelled signal, but a "less vivid" signal (that is, less vivid than how they could be felt were they not predicted at all). This way of using the notion follows Frith's own usage: see, e.g., Frith (2005; 171): "Because we can predict the sensory consequences of our own actions, our response to these sensations is attenuated. A touch we apply to our selves feels far less intense than the same touch applied by someone else".

would therefore be indistinguishable from externally generated sensations, hence *causing* a confusion between the self and the other" (Blakemore 2003; 653, my italics). It seems that what is proposed here is that the lack of attenuation plays a role in generating the experience of passivity. Thus, attenuation would be necessary in order to experience agency, at least because non-attenuation generates an experience of passivity.

However, we may wonder whether some theorists would maintain that, just as the lack of attenuation plays a role in causing experiences of passivity, attenuation is involved in causing agentive experiences. The two quotes above leave open the question of whether attenuation plays some causal role in generating the feeling of agency. It could be that a lack of attenuation generates an experience of passivity while attenuation does not generate an agentive one (because, say, the experience of agency is generated by default—only when signals are not attenuated is it altered). The following quote from Frith (2012; 53) suggests that his position is (or at least has been) that the sense of agency is an effect (at least partially) of attenuation[2]: "Studies of binding have shown that the experience of agency *depends* upon two components, first an advance prediction of the effects of the action and, second, a retrospective component whereby the nature of the outcome changes the experience of the action" (my italics). Frith is talking here about intentional binding (see below), not attenuation[3]. However, it is possible to speculate that if he now thinks in these terms (dependency) about intentional binding and the sense of

[2] Attenuation can at most be a partial cause (a causal factor) of the sense of agency, for attenuation by itself does not cause the feeling of agency—attenuation quite commonly occurs in response to expected events in general. I owe this specific observation, as well as the general reflection on the role of attenuation in the etiology of the sense of agency, to an anonymous referee.

[3] This seems to attest to a change in Frith's mind concerning the etiology of the experience of agency; but I will not dwell on it, for it would only add confusion.

agency, he may also think, or have thought, in similar terms about attenuation and the sense of agency[4]. My impression is that comparator theorists have been unclear on this issue. However, I do think that, as I say above, most would agree that attenuation is a necessary condition for experiencing agency (that is, a subject cannot experience agency if the incoming feedback signal is not attenuated).

Recently, another candidate for being connected in a special way with the sense of agency has been proposed. Haggard and colleagues (see, e.g. Haggard and Clark, 2003) have discovered a phenomenon they call 'intentional binding'. The phenomenon consists of perceiving the feedback from an action closer in time to its predicted effect than it actually occurs. This phenomenon consists in perceiving the action and its predicted effect closer in time than they actually are. On the one hand, the action is "moved forward", experiencing it as taking place later in time. On the other, the effect is moved backwards, as it is experienced as occurring before it actually occurs. The dislocation of the action is claimed to occur ante hoc and to be due to predictive mechanisms; while the displacement of the effect is said to be post hoc. Now, Haggard's claim is that intentional binding is correlated with the sense of agency and that schizophrenia patients present some disturbances in binding, as they tend not to move the action forward. As noted above, Frith (2012) apparently thinks more in terms of dependency than merely in terms of correlation.

[4] In fact, it makes more sense to think that the sense of agency depends on attenuation than it does to think that it depends on intentional binding. Attenuation allegedly helps distinguish what has been done by the subject from what the world has done. Intentional binding does not appear to play such a role.

As can be seen, both attenuation and intentional binding involve the predictive system working well, and studies reveal that schizophrenia patients present deficits in both functions. Thus, it seems plausible that there is something malfunctioning in the predictive system of schizophrenia patients and that the malfunction has something to do with some of the passivity phenomena they may experience. The question now is whether *all* passivity phenomena can be explained in terms of failures in the predictive system. We can grant that delusions of control, i.e., the feeling and firm belief that one's movements are controlled by someone else, can be explained as malfunctions of the predictive system (either because it fails to make predictions or because it makes too many—see Hauser et al., 2011). What about thought insertion and AVHs though; does the comparator model provide a simple explanation of them?

The comparator model in the cognitive domain

Many psychologists see the transition from the body to the mind as utterly unproblematic. It is unproblematic because it is simply assumed that thinking is an action (see Frith, 2012, where he seems to wonder for the first time whether this is so). For instance, Ford et al. (2007) approvingly cite Jackson's (1958) claim that "thinking is our most complex motor act" (p. 459). It is noteworthy, and somewhat paradoxical, that some philosophers strive to show that thinking is an action by resorting to the comparator model (see Peacocke, 2007) and by invoking the possibility that it explains mental passivity phenomena; whereas some psychologists think the model works in the mental realm because they assume that thinking is a mental action.

However, the explanation of overt actions (and of the mechanisms of their self-ascription together with the breakdown of those mechanisms) cannot easily be exported to the cognitive domain; in particular, the same explanation would not seem to account for thought insertion and alien voices. The explanation of thought insertion involves thinking being caused by previous intentions, and thoughts being compared to predictions generated by forward models. If all goes well, thoughts are self-ascribed—i.e. the subject feels that *he* is thinking. Otherwise, just as in the case of delusions of control, where the subject thinks someone else is making him act, the subject is led to think that someone else is putting thoughts into his mind. The theory offers the following explanation for alien voices (AVHs): we very often engage in inner speech. In order to talk to ourselves, we have to form the intention to do so, give precise instructions about how to satisfy that intention and check whether our intention is satisfied. When things go well, we generate the typical inner speech which we self-ascribe. However, if the monitoring mechanism breaks down, our inner talk is attributed to someone else. *Prima facie*, the model can even explain very elegantly why alien voices are voices while inner speech does not feel like a voice (it lacks pitch, volume and intonation, see Vicente and Martinez-Manrique 2011): while inner speech perceived as such is attenuated because it accords well enough with the predictions, voices result from the non-attenuation of the reafferent signal[5].

[5] I will revisit and question this apparently neat explanation below. Please note that inner speech is here said to be attenuated in the sense of being less vivid than voices (see fn. 1). Many authors seem to endorse the ideas that (a) inner speech is an attenuated signal, and (b) that it is attenuated because it is adequately predicted. For instance, Ford and Mathalon (2005; 183) write: "ERP evidence from the experiment described above and others in this series … suggests that auditory cortical responsiveness is reduced during talking and inner speech. We assume this is due to a corollary discharge from frontal speech production areas to speech reception areas in the temporal lobe". I assume this is the "official" explanation that the comparator theory offers for the distinctive phenomenology of inner speech.

So, it seems as though the comparator model has great explanatory power and that it could be extended smoothly into the mental realm and deployed to explain the two most relevant schizophrenia symptoms there. Yet, there are a number of issues that generate some discomfort when the theory is applied to the cognitive domain; in particular, to the explanation of thought insertion (I reserve discussion of AVHs for later sections). First of all, as Stephens and Graham (2000) hold, the account cannot explain the difference between inserted thoughts and thoughts that we feel we have not generated (unbidden thoughts). Next, there is an issue concerning functionality: comparisons seem necessary in perception in order to distinguish when I am moving as opposed to when the world is moving; and in the motor system in order to discern intended from unintended movements. However, what is the purpose of having a monitoring system in the case of conscious thinking? It seems that there is no distinction we have to track here between what we are doing and what the world is doing to us. Moreover, as Gallagher (2004) points out, why should we have to have a monitoring system for this kind for *conscious* thought, i.e., mental events that are susceptible to conscious control?[6]

However, the most pressing problem for the comparator approach to thought insertion is that it is a mystery exactly what is compared to what. A minor concern here is that it is not clear how we compare the conscious thought to the output of the forward model, which presumably is unconscious (see Campbell, 1999). The major concern, however, is that it is not clear what the efference copy and the prediction delivered by the forward models consist of. In the case of bodily actions, the efference copy is of a motor command and the prediction consists of sensory and proprioceptive information. What

[6] One possible response is that not every trait has to be adaptive.

are the equivalents of motor commands and somatosensory predictions in the case of mental actions such as judgments? That is, how does the model really work?

Some authors (see, e.g.,Carruthers, 2011) hold that, as a matter of fact, it does not make sense to apply the model to thought insertion; only motor commands trigger efference copies, and there are no motor commands involved in thinking. Vosgerau and Newen (2007) are more detailed in their critique: the model requires that the comparison process is triggered by an intention which contains a representation of the desired outcome, which results in a command. That command then triggers some processes that result in movements. The comparator mechanism is in charge not only of comparing predictions with results, but also of monitoring the process—and of calling for adjustments when required. Now, when we try to think about tokening thoughts in these terms, we mostly find mismatches. To begin with, intentions can be considered to be thoughts; at any rate, they are "mental actions"—and the model for tokening thoughts applies to all alleged mental actions. However, if intentions are thoughts or mental actions, they will require previous intentions to form them, which means that we face an infinite regress. Secondly, tokening a thought in consciousness does not seem to be a process that can be adjusted or monitored. The intention to form a thought does not contain a blueprint of an action which is implemented by successive motor processes. So the idea that the comparator mechanism keeps track of how the instructions delivered by the intention are developed does not make sense.

As has been said, some psychologists, including Frith, have simply assumed that thinking is a complex motor act. However, it is certainly difficult to explain the tokening of a

thought even in the general terms we use to explain motor acts. It makes doubtful sense to speak of commands, processes and consequent sensory effects in the case of having a thought. In short, we seem to have no idea of what (if it is not a prediction issued by an efference copy of a motor command, then what is it?) is compared to what (if it is not incoming sensory signals, then what is it?).

Some authors have attempted to explain how the model could work in the cognitive domain. For instance, Proust (2006) makes use of the comparator model to account for the sense of agency associated with so-called “mental actions”. However, it seems that not all mental actions are amenable to the kind of explanation she proposes. Proust uses the examples of trying to retrieve a name from memory and of planning an action of a type that has been performed before. Now, it seems that both cases involve processes, and results which can be predicted. In the first case, the predicted state is the state of knowing—or the state of not knowing. In the second case, according to Proust, the subject may predict what the reafferent signal is going to be like based on prior experience with similar tasks. The question is whether the case of thought (and of thought insertion) is similar to these two examples. The intuition is that, because of what is explained in the previous paragraph (and developed at length by Vosgerau and Newen, 2007), it is not. Frith (2012) echoes Vosgerau and Newen’s worries approvingly, while putting them in somewhat different terms: entertaining a thought, he says, involves neither kinematics nor sensations, which seem to be necessary ingredients of the comparator mechanism. I think that this way of putting things summarizes the problem nicely.

Thought insertion and AVHs

The criticisms above are not knock-down arguments, but they reveal that there is indeed a puzzle concerning thought insertion. Now I want to turn to possible answers to these worries. I will examine what I take to be the most promising general take on the issue and distinguish two alternative paths within that general approach. The general approach consists of claiming that (conscious) thoughts are not "pure thoughts"[7], but that they are always carried by some perceptual vehicle; in particular, by inner speech. Once we think about thoughts in this way, e.g., as contents of strings of inner speech, then it seems that it makes sense to think that there are motor commands, reafferent signals and so on. That is to say, thinking resembles acting in the key respects necessary for the comparator model to be applicable. In other words: if thought insertion is misattributed inner speech, we seem to have a clear idea of what is compared to what and how the comparison takes place. The command is to produce a string of words, and the reafferent signal is the string of words produced. Allegedly, a prediction is issued regarding what the string of words should be like, and one can compare that prediction to the outcome. Nonetheless, it is possible to speculate as to why inner speech is checked for authorship: basically, if outer speech is checked for authorship—because all overt behavior is so checked—it is reasonable to expect that inner speech, which involves many of the same systems as those involved in outer speech, will be also checked, even if there is no point in doing so. That is, the comparator system for inner speech rides piggy-back on the comparator system for outer speech.

[7] Hurlburt (2009) calls them 'unsymbolized thinking'.

This general approach seems, on the other hand, *prima facie* perfectly reasonable. The identification of inserted thoughts with misattributed inner speech is independently plausible and does not require that one endorses the comparator theory. That is, it is possible to resort to pieces of evidence, theories and arguments that are independent of the comparator theory. For instance, a good number of comparator theorists hold that the function of inner speech is to make thoughts available to the consciousness. Carruthers (1996) defends the strong thesis that inner speech is the vehicle of conscious thinking. In Carruthers (2006), he weakens his former claim, but he still proposes that we are only conscious of thoughts that are delivered via a perceptual vehicle; inner speech being the most fundamental. This hypothesis is shared by other authors (see, e.g. Jackendoff, 1996, 2012). It can also be held that not all that is conscious is perceptually conscious, but that we do recruit perceptual vehicles—and especially inner speech—to raise thoughts to consciousness (Martínez-Manrique and Vicente, 2010). If this hypothesis were right, that is, if we mostly used inner speech to bring thoughts to consciousness, it would be *prima facie* reasonable to speculate that thought insertion consists of the misattribution of strings of inner speech[8].

Now, the first way to flesh out this general approach consists of identifying thought insertion with AVHs (for defenses of this identification, see Graham and Stephens, 2000, and Langlad-Hassan, 2008). As I have said, the comparator model seems capable of covering AVHs: what goes wrong in AVHs is that inner speech is not recognized as such due to a breakdown in internal monitoring. The production of inner speech involves having the intention to form a string of inner speech by activating the linguistic system.

[8] I revise the cogency of this inference below.

The instructions delivered to the linguistic system are copied and sent to the forward models, which in turn generate and deliver a prediction about the expected sensory feedback. When the silent talk is finally produced, the feedback from it is compared to the prediction and self-ascribed if they match. If things go wrong because the efference copies are not correctly delivered or because the forward models do not work or make the wrong predictions, inner speech is not felt as one's own and is attributed to an external agent. Moreover, and crucially, the incoming signal is not attenuated: instead of being experienced as inner speech, i.e., silent words with no tone, pitch or intonation, it is experienced as a voice.

In the next section I aim to show that this rendering of the explanation of thought insertion may be problematic because the comparator account of AVHs is itself problematic. Here I first present some preliminary criticisms of the proposed reduction of thought insertion to AVHs. The main criticism is the following: it might be that patients may mistake alien thoughts for alien voices expressing those thoughts. However, patients who suffer from both AVHs and thought insertion repeatedly claim that they are different phenomena (or, at any rate, that they can be distinguished on the basis of their phenomenology)[9]. More importantly perhaps: they are treated as different phenomena in

[9] I was unable to access transcripts of interviews with patients, but I passed out a questionnaire to some psychiatrists from the local hospital. To my question regarding whether patients who suffer both AVHs and thought insertion differentiate between the two, I got this response: "they usually differentiate them, although sometimes it is complex" (reported by Margarita Saenz, Psychiatry, Santiago Hospital, Vitoria-Gasteiz). See also the following Internet forums:
http://www.schizophrenia.com:8080/jiveforums/thread.jspa?threadID=31, and
http://www.schizophrenia.com:8080/jiveforums/thread.jspa?messageID=260048.

the clinical literature. This, I think, is enough to cast serious doubt on their identification[10].

Then there is a point I want to mention concerning the theories that I have said would provide independent support for the explanation of thought insertion as misattributed inner speech in any of its versions—thought insertion as AVHs or thought insertion as simply misplaced inner speech[11]. This point does not directly tell against the explanation, but deprives it of some of its support. The issue is that it is certainly possible that the perceptual theory of consciousness, as well as the weaker thesis that we have to recruit perceptual mechanisms (in what concerns us here, inner speech) to raise thoughts to consciousness, is wrong. Hurlburt (see, e.g. Hurlburt and Akhther, 2008) tells us that his experimental subjects report having bare thoughts, that is, thoughts not accompanied by inner speech or images. In a discussion with Carruthers, Hurlburt (2009) concedes that perhaps these thoughts are not totally devoid of sensation-like phenomenology, but this concession does not change things much; the minimal sensations or feelings Hurlburt speaks of, even if they are perception-like, are only accompanying the thoughts: they cannot be said to be carrying or standing in a content relation to them. The fact that there is this kind of conscious thinking independent of any perceptual vehicle does not imply that inserted thoughts are misattributed strings of inner speech; it might be that thought insertion only occurs when thoughts are put into words. However, Hurlburt's data seem to imply that thinking can be pursued by autonomous means, and so some conscious

[10] As an anonymous referee pointed out, this distinction might be something like a cultural construction. Barrett (2004) reports having found nothing like thought insertion among the Iban in Australia, who seem to think about thinking in a different way. More cross-cultural research is needed to address this question.
[11] The difference between "thought insertion as AVHs" and "thought insertion as misplaced inner speech" is that under the former hypothesis, thought insertion would be inner speech which sounds like a voice, while under the latter it would be inner speech that, while it is not self-attributed, still sounds like inner speech.

thoughts do not involve inner speech. That is, the data can be accommodated by the explanation of inserted thoughts as inner speech as a matter of principle, but they cannot be accommodated by the more general theory that motivates such an identification.

Are AVHs misattributed inner speech?

I take it that the identification of AVHs with thought insertion is itself problematic. However, I now want to explore another route that would not only undermine the reduction of thought insertion to AVHs, but would question whether it is even advisable to attempt such a reduction. I now suggest that there is no good extant comparator model explanation of AVHs in terms of misattribution of inner speech[12]. This would have several consequences. First, and obviously, the preferred comparator account of AVHs may fail—and the theory appears to lose part of its unificatory power. Second, the hypothesis that thought insertion is misplaced inner speech would be problematic *insofar as thought insertion is identified with AVHs*. Third and finally, the comparator model of thought insertion would be problematic *insofar as thought insertion is identified with AVHs*. To this end, I will make use of a study conducted by Langdon and colleagues (see Langdon et al., 2008), which relies on reports made by patients who suffer from AVHs and compares them to reports made by healthy individuals. The issues the study focuses on are: first, whether patients with AVHs have less inner speech than control subjects, and, in particular, whether they have less *expanded* inner speech; second, whether the

[12] In the final section I present other criticisms of the comparator explanation of AVHs, based on the notion of attenuation. To advance: I do not think that the attenuated character of inner speech is due to its being successfully predicted. This means that the alleged lack of attenuation of AVHs is not due to unsuccessful predictions, and so the comparator theorist does not have a ready explanation of why AVHs are experienced as voices.

pragmatics of AVHs are similar to the pragmatics of inner speech, focusing in particular on the way AVHs and inner speech address the subject; and third, whether the phenomenological qualities of AVHs (vocal characteristics—such as perceived gender—speed, form and volume) resemble the qualities of inner speech. I will explain.

It seems to follow from the comparator approach to AVHs that patients who experience AVHs should engage less often in inner speech, as some tokens of inner speech are misattributed and turned into hallucinatory voices (this hypothesis is explicitly endorsed by Lysaker and Lysaker, 2005). Fernyhough (2004) proposes that it will be easier to find misattribution in the case of expanded inner speech, because, according to his Vygotskyan theory, inner speech (which is usually condensed or fragmentary) typically takes an expanded form when the subject confronts a cognitively demanding task or is under stress. Thus, the first thing the study aimed to test is whether these consequences of the Frithean model hold.

Another apparent consequence of the Frithean account of AVHs is that there should be no difference in the pragmatics of AVHs *vis à vis* inner speech. That is, if the alien voices typically address the subject in the second or third person (i.e., they use the pronouns ‘you’ and ‘your’ or ‘he/she’ and ‘his/her’ etc.), then so should inner speech. The same holds for the typical speech acts alien voices and inner speech perform: if the alien voices typically issue commands, inner speech should adopt the imperative mood as well. The theory does not offer any explanation for why there should be a change in the way the subject feels addressed. If AVHs are misplaced inner speech, AVHs should address the subject in the same way as inner speech does.

Lastly, it also seems to follow from the Frithean model that inner speech and AVHs should share some phenomenological qualities. It seems difficult to explain why misattributed inner speech should be experienced as, e.g., a female voice in the case of a male subject, or as slower, louder or more elaborate.

It turns out that all these putative consequences of the Frithean model were falsified by the study. Schizophrenia–AVH patients did not differ significantly from controls in the frequency of their inner speech. In particular, they did not differ in the frequency of their expanded inner speech. The patients did not experience their inner speech as resembling voices—rather than silent words—and it was not gendered—while AVHs were gendered and clearly identified as voices; and as recognizable voices. There was no correlation in general between the relevant features of inner speech and AVHs. In particular, there was no correlation with respect to: (a) the use of the patient's own name, (b) the use of second- or the third-person pronouns, (c) speed, volume and intelligibility, or (d) hearing voices conversing/hearing voices talking directly to oneself and experiencing inner speech as a kind of conversation/experiencing it as self-addressed. Also, there was no difference in any of these regards between patients with a high frequency of AVHs and those with a low frequency of AVHs. It is also significant that patients reported answering their AVHs in inner speech; for, according to the authors of the report, it seemed odd that the responses to alleged misattributed inner speech should not be similarly misattributed.

Langdon et al. conclude from their research: "if inner speech, conceived as the act of internal self-talk, is the raw material of all AVHs, then there should be similarities between the phenomenological characteristics of patients' verbal thought and their AVHs. We found no evidence to support this prediction" (2009; 662). I think that this conclusion

is well motivated, although there are aspects of the study that deserve some criticism. Above all, I think Langdon et al. are wrong in assuming that there should not be a difference in the phenomenology of AVHs, which are experienced as voices, and that of inner speech. In fact, one of the virtues of the comparator model is that, in principle, it can explain why inner speech is not felt as a voice, while misattributed inner speech should sound more like a voice: while inner speech is an attenuated signal, misattributed inner speech is not attenuated (I come back to this issue below). That said, however, it does seem mysterious that voices should be gendered, and in particular that male patients could experience female voices[13]. Now, that issue aside, it does seem that the rest of the study reveals facts that cannot be predicted (or explained) using the comparator model. This, by itself, does not refute the approach—a false prediction does not refute a theory: theories go on living in spite of false predictions, which are then treated as puzzles—but it certainly casts doubts on the comparator explanation of AVHs as misplaced inner speech. The differences between inner speech and AVHs seem to be substantial enough to rule out the possibility of the model accounting for them in a straightforward fashion. Indeed, they make it look as though AVHs *are not* misattributed inner speech.

Thought insertion

Now, where does this leave us? The evidence seems to undermine the hypothesis that AVHs are misattributed inner speech, which means that the preferred comparator account of AVHs may fail, and so, on the face of it, the comparator theory loses part of its

[13] The hypothesis discussed here is that AVHs are misattributed inner speech. As inner speech seems to be a faint copy of one's own voice, it is strange that, when misattributed, it should sound like a female voice if the subject is male. An alternative hypothesis about AVHs, suggested by a referee, is that they are misattributed verbal imagery—not inner speech. This is an interesting hypothesis. In fact, I think it is probably a better hypothesis, all things considered. I will not pursue it here, however, as this is not the place to discuss verbal imagery which is not inner speech.

unificatory power. The evidence further undermines the hypothesis that thought insertion is misplaced inner speech insofar as thought insertion is identified with AVHs. However, the comparator theorist is not forced to identify thought insertion with AVHs at all. I said that treating thought insertion as AVHs had an advantage, specifically, that it provides a clearer idea of what is compared to what in terms of thoughts (as well as unifying two apparently disparate symptoms). However the comparator theorist can now follow one of two paths: in the first place, stick to the original idea that thought insertion is a breakdown in the capacity to monitor not inner speech, but thought processes; or secondly, rework the thesis that thought insertion is misattributed inner speech. I will explain.

The study reveals that AVHs are not misattributed inner speech. However, it does not reveal that *thought insertion* is not misattributed inner speech. If thought insertion is not the same thing as AVHs, then perhaps AVHs are not misplaced inner speech but thought insertion is. That is, perhaps the trick is precisely to keep AVHs and thought insertion apart. If these two symptoms are kept apart (as it is customary to do), then it seems possible to maintain that the experience of thought insertion results from defective monitoring of inner speech. For instance, it could be claimed that subjects issue instructions to tell themselves things, those commands are followed, and the subjects engage in “normal” inner speech; that is, self-talk which is silent, perhaps condensed, gender-neutral and pragmatically non-deviant. However, the monitoring process goes wrong, and the subjects are incapable of recognizing the inner speech as their own. As a result, they refuse to self-ascribe it and ascribe it to an external source. Indeed, this kind of account squares nicely with reports from patients who make claims such as: “thought

insertion to me is when they put words in my head"[14]. Here the subject is not claiming to hear voices, only to be aware of some words that are not self generated.

So, the defender of the comparator model can follow one of two paths: the first consists of maintaining that thought insertion results from a misattribution of the pure thoughts Hurlburt speaks of; the second is to claim that thought insertion stems from a misattribution of the self-generated perceptual vehicles we use to think (inner speech being the most common one). It is even possible to adopt a mixture of the two and claim that thought insertion is produced whenever pure thoughts, inner speech or images are misattributed.

However, do any of these responses to the challenge show much real promise? It seems that they do not. To begin with the most obvious: resorting to pure thoughts is clearly a dead end. If what is claimed to be misattributed are bare thoughts—not inner speech or images—we are back to the starting point: it is difficult to make sense of the idea that there are initiating commands, efference copies and predictions which are compared to outcomes of something that has no sensory realization. The move is not advisable at all. Now, what about the idea that thought insertion results from the misattribution of inner speech? The proposal sounds much more promising. Here there is an intention to produce silent speech, so there is an initiating command, and the result of that intention can be compared, via the production of an efference copy of the command, to a prediction issued by a forward model. Moreover, this proposal seems to respect the phenomenology

[14] Retrieved from:http://www.schizophrenia.com:8080/jiveforums/message.jspa?messageID=49; post by "Toast". Accessed 02/08/2013.

reported by some patients who feel inserted thoughts not as voices but as silent words put into their minds.

However, this proposal has one main problem and has to face strong challenges. Let us begin with the challenges, some of which also affect the proposal that thought insertion is misattributed bare thoughts. To see what these challenges amount to, it is possible to make use of the dialectics employed by Langdon et al. (2008) and to test whether some of the phenomena surrounding thought insertion are predicted and explainable by this particular instantiation of the comparator model. The following points are those which I regard as problematic in this sense.

First: it is very often the case that inserted thoughts do not make sense at all. A typical thought that a subject may entertain is that an invasion of the Earth is imminent, or that he should kill God (see Frith, 1992). It is not plausible to think that such thoughts feel "alien" just because the monitoring mechanism has gone wrong. If the comparator model were right, we should observe "normal" thoughts that the subject refuses to self-ascribe and some "weird" thoughts that the subject does self-ascribe. That is, there should be no difference at the level of content-type between self-ascribed thoughts and inserted thoughts. However, there seems to be a big difference[15].

[15] See, e.g. Jackson & Fulford (1997) who cite the following report from a patient: "the things that come are not the things that I have been thinking about... They kind of short circuit the brain, and bring their message". It seems that anyone with personal or clinical knowledge of thought insertion will confirm that the content of inserted thoughts is peculiar. A response to the questionnaire that I passed out to local psychiatrists reads: "Usually, in this kind of phenomena of mental automatism, content is extravagant and bizarre; it's not "like going to buy some bread".

Second: inserted thoughts are usually intrusions that divert subjects from what they are focusing on. Again, there is no reason why this should happen if inserted thoughts are just thoughts which lack adequate monitoring.

Third: there is a problem with what Langdon et al. termed "the pragmatics" of inserted thoughts. We have seen that their study reveals that AVHs differ from inner speech, among other ways, in questions such as how they address the subject. I have found no equivalent study that compares inserted thoughts to normal thoughts. Suppose, however, that inserted thoughts were misplaced strings of inner speech. We should then expect that inserted thoughts would not differ from the usual inner speech in, for instance, the mode of presentation of the subject. That is, if subjects tend to speak to themselves in terms of 'I', their allegedly misattributed inner speech should also use first-person pronouns. However, this is not usually the case. Moreover, inserted thoughts often take the form of threats, warnings, commands or admonitions; speech acts that do not occur frequently in inner speech. That is, inserted thoughts appear to be very different from the normal flow of inner speech and of thinking in terms of content, form and pragmatics[16].

It may be that some of these problems do not exclusively affect the explanation of thought insertion we are considering. For instance, it is often said that the comparator approach provides a good explanation of anarchic hand syndrome (though see Pacherie, 2007). However, the alien hand usually does very weird things, such as picking food up from someone else's plate or giving an injection to the doctor who is attending the hand's owner. It seems that, just as in the case of utilization behavior, subjects—or the subject's

[16] See, e.g. Moritz and Larøi (2008), where it is said that "[inserted thoughts] are frequently experienced as ego dystonic (i.e., are experienced as inconsistent with the person's belief system) and … share similarities in form, content and triggers (e.g., stressful events)".

hand, in this case—do just about anything that can possibly be done, no matter how appropriate the action is. However, if the hand feels alien because its behavior is not predicted, there is apparently no reason why its typical behavior should be so odd. Nonetheless, it does seem as if part of the syndrome consists of the hand typically doing weird things. It might be suggested that what we have in this case is utilization behavior plus a breakdown in the monitoring of the anarchic hand; but it is still odd that the utilization behavior should only be displayed by the anarchic hand[17].

Now, let us move to the problem that affects the proposal that thought insertion is simply misattributed inner speech; that is, it is inner speech which "sounds" just like normal inner speech but is attributed to an external source. The problem is this: the comparator approach, as we have seen, attributes a prominent role to the attenuation of the reafferent signal. According to what we have seen is claimed by representative defenders of the theory, a lack of attenuation is the basis of the experience of alienation, and only non-attenuated signals generate experiences of passivity. However, according to the present

[17] A further problem which seems to generalize to other passivity phenomena—and that is why I only mention it here in a footnote—is how the comparator model explains other-attribution. Patients do not simply refuse to self-attribute their thoughts, but often they attribute them to some particular entity, say, the FBI, aliens, or whatever. It seems to be at least an open question whether this kind of particular other-attribution is part of what has to be explained in thought insertion. Philosophers such as Hoerl (2001), as well as psychologists such as Heinks-Maldonado et al. (2007), hold that the proper *explanandum* of thought insertion is just the failure to self-attribute. One reason for claiming this is that some patients exhibit only this negative symptom; that is, they maintain that some of their thoughts are not their own but they do not take the further step of attributing them to a particular alien entity. However, many other patients do take that further step. What is an open question here is whether this step is part of the phenomenon of thought insertion. For instance, it could be argued that such other-attribution is produced by some sort of inference to the best explanation, detached from the phenomenon itself: if I am not the one thinking my thoughts, then some very powerful entity must be placing them inside me. Yet, many patients do usually understand that the best explanation for their inserted thoughts is that they are the product of a malfunction in their brains: the problem is that their inserted thoughts do not feel that way; rather they feel like thoughts inserted by the FBI, aliens, etc. Jeannerod (2003) proposes a simulationist model which in principle can account for this. The model includes a "who" system in charge of distinguishing "me" from "another agent" (instead of simply distinguishing "me" from "not me"). That model takes the phenomenon of other-attribution seriously and it can probably account for it satisfactorily. However, we have yet to see how the simulationist model applies in the cognitive domain.

proposal, the incoming signal seems to be attenuated. Inner speech is felt as inner speech, not as a voice, so we should assume that it has been attenuated. If it has been attenuated, however, that is because it has been adequately predicted; moreover, if it has been attenuated, it has been marked as endogenously produced. So it should not be perceived as alien.

Now, it is possible to wonder whether the signal has actually been attenuated: why should we believe that inner speech is an attenuated reafferent signal? The reason is quite simply that comparator theorists say it is (see, e.g. Frith, 1987, Ford and Mathalon, 2005, Ford *et al.*, 2007) and that it makes sense to think that, according to the theory, inner speech is attenuated speech. The impact of claiming otherwise would be considerable: first, the preferred explanation of AVHs would have to be abandoned; we do not seem to be able to explain the "voice experience" in terms of a lack of attenuation of inner speech that is usually attenuated. Secondly, we would also have to drop the explanation of why inner speech is felt as it is felt (i.e., relatively poor in phenomenology): we could not claim that it is phenomenologically scarce because it is an attenuated signal. Finally, the theory seems to require that for something to be misattributed, there has to be a lack of attenuation (or at least, it requires that, if there is a reafferent signal, that signal has to be non-attenuated in order to be misattributed)[18]. So, suppose that inner speech is not an attenuated signal; then, what is *not* attenuated in thought insertion?

There are possible ways to circumvent this problem, but they do not help to evade the dilemma that is now emerging: either we account for AVHs or we account for thought

[18] At any rate, it requires that: if there is a reafferent signal, then that signal has to be non-attenuated if it is misattributed.

insertion. I now explain how the problem can be circumvented before then moving on to develop the dilemma in the closing remarks of this section.

Inner speech is phenomenologically poorer than voices: it lacks volume, pitch, tone, candescence, etc. One way to explain this difference in phenomenology is, as we have seen, by maintaining that inner speech is a signal that, because it is adequately predicted, is attenuated. However, there are alternative explanations. For instance, Carruthers (2006) holds that inner speech is quasi-produced speech, and that all quasi-produced actions are phenomenologically poor. According to his account, which draws on Jeannerod's (see, e.g. Jeannerod, 2003), a quasi-produced action results from issuing motor commands which send efference copies but are then aborted. The forward models work with the efference copy and yield a prediction that can produce an image-schema of the action, that is, an image that is phenomenologically poor. So, the reason why inner speech feels the way it does is because it is an image-schema of a speech act. In general, it seems that endogenously produced and consumed signals (call them 'private' signals) are attenuated with respect to signals of the same relevant kind that arrive via the senses (in Hume's terms, they are less vivid). This holds for images and inner speech as well as for emotions, feelings, etc.

Now, the consequence of substituting the original comparator explanation of the phenomenological properties of inner speech for this one would be that the comparator model requires some adjustments when it is exported to the cognitive domain. In particular, it has to be stripped of the idea that signals are attenuated as a result of successful comparison. Now, signals are already attenuated before the comparison takes place. I do not know whether abandoning this idea is going to create havoc in the general

approach or not; that depends on what role comparator theorists ultimately assign to attenuation. If they want to maintain that attenuation is in all cases necessary for the sense of agency, then the problem is going to be serious. However, it may be that theorists end up thinking that the role of attenuation in the theory has been overrated by some authors or in some cases.

Nonetheless, there is another, less direct and more damaging consequence of forsaking the idea that inner speech is an attenuated signal because it is adequately predicted: the comparator theorists have to abandon their current explanation of AVHs as well. The current explanation of AVHs tells us that voices are experienced as voices because inner speech is not attenuated. We have seen already that this cannot be the whole story, for if a male's inner speech was not attenuated it should never be heard as a female voice. Leaving that aside, the point is now: why do alien voices sound like voices? The comparator theorist does not have an answer to this question in terms of forward models, predictions and comparisons. That is to say, the comparator theory does not seem to be able to explain AVHs (at least, not if they insist in adopting the "inner speech" approach). And this takes us to the dilemma: it seems that, if we establish that AVHs and thought insertion are different phenomena, the comparator theory has to choose between explaining one and explaining the other.

The comparator theorist can choose to explain AVHs. This would lead to a defense of AVHs as being experienced the way they are experienced because inner speech is not adequately predicted and thus is not attenuated[19]. But then there is no possible

[19] Perhaps I should insist that I do not think that this is the best explanation of the attenuated character of private signals. The view that they are simulated, or quasi produced, seems to be much more promising.

explanation of why inner speech in thought insertion—that is, if thought insertion is in effect misattributed inner speech—is not experienced as a voice. In contrast, if the comparator theorist chooses to explain thought insertion as misattributed inner speech, then the preferred explanation of AVHs must be sacrificed: in thought insertion, inner speech is misattributed, but experienced the way inner speech is usually experienced; which means that AVHs are some thing else, not misattributed inner speech. This dilemma does not arise if thought insertion and AVHs are ultimately the same phenomenon. However, they seem to be different phenomena. At any rate, they are *perceived* as different phenomena, and that is enough to motivate the dilemma, given that what we aim to account for is the *phenomenology* of AVHs and that of thought insertion.

Conclusion

I propose to sum up the main points of this paper as a set of open questions that remain for the comparator theory:

1. What is compared to what in thought insertion?

2. Why do AVHs have the content and pragmatic and phenomenological characteristics that they have?

3. Why do inserted thoughts have the content and the pragmatic characteristics that they have?

4. Why are “private” signals attenuated?

5. Why are AVHs experienced as voices?

6. If inserted thoughts are misattributed strings of inner speech, why are they experienced as *silent* strings of inner speech?

This paper has explored possible answers to some of these questions and I have tried to extract the consequences of responding to them in one way or another. To this end, I started with the "What is compared to what?" problem of thought insertion (question 1). I then suggested a general approach to tackling that question: assume that thinking is using inner speech, and explain thought insertion as misattributed inner speech. After that, I distinguished two ways of implementing that idea: (i) identify thought insertion with AVHs, and (ii) regard thought insertion as *only* normal inner speech that is misattributed. In both cases, we could observe the same kind of problem: both AVHs and thought insertion have intriguing and peculiar properties that the comparator model does not predict (questions 2 and 3). This casts doubt on the comparator explanation of thought insertion (whether it is reducible to AVHs or not) *and also* on the very explanation of AVHs. Finally, we arrived at the issue of attenuation (questions 4, 5, and 6). Here it emerged that the preferred comparator approach to attenuation could explain AVHs but not thought insertion; but that what probably amounts to a better view of attenuation would be unable to explain AVHs. In short, the final stage along the path that began with the "What is compared to what?" problem of thought insertion is: maybe the real problem is AVHs.

More cautiously, I suggest that it can be concluded from the discussion here that no answer to the six questions above is free from problems, and so those six questions do not have convincing answers to date. Yet, comparator theorists must surely tackle them if

they want to deliver an explanation of passivity phenomena in the cognitive domain; that is, if they want to deliver the unifying explanation that the theory initially promised.

Acknowledgements

I want to express my gratitude to the Logos group in Barcelona (particularly to Manuel García-Carpintero and to Pepa Toribio), as well as to my colleagues at Granada, and especially to Fernando Martínez-Manrique. I am very grateful to Margarita Saenz and Sonia Ruiz de Azua from the Psychiatry Unit at the Santiago Hospital, in Vitoria. They did help me a lot. Finally, I want to thank Toffa Evans for the English revision. Research for this paper was funded by the Spanish Government, research projects FFI2010-15717 and FFI2011-30074-C02-02.

References

Barrett, R. J. (2004). Kurt Schneider in Borneo: Do First Rank Symptoms Apply to the Iban? Schizophrenia, In J. H. Jenkins and R. J. Barrett. *Culture and Subjectivity: the Edge of Exp*erience.

Blakemore, S-J. (2003) Deluding the Motor System. *Consciousness and Cognition* 12: 647–655

Blakemore, S-J. and Frith, C. (2003) Self-awareness and Action. *Current Opinion in Neurobiology* 13:219–224

Blakemore, S-J., Wolpert, D. and Frith, C. (2000). Why Can't You Tickle Yourself?. *NeuroReport*, 11:11-16.

Campbell, J. (1999). Schizophrenia, the Space of Reasons and Thinking as a Motor Process, *The Monist* 82: 609-625

Carruthers, G. (20121) The case for the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and Cognition* 21: 30-45

Carruthers, P. (1996). *Language, Thought, and Consciousness*. Cambridge University Press.

Carruthers, P. (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought.* Oxford: Clarendon Press.

Carruthers, P. (2011). Action-Awareness and the Active Mind. *Philosophical Papers* 38 (2):133-156.

Feinberg, I. (1978). Efference copy and corollary discharge: Implications for thinking and its disorders. *Schizophrenia Bulletin* 4: 636–640.

Fernyhough C (2004). Alien voices and inner dialogue : towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology* 22: 49–68.

Ford, J. M., Roach, B. J., Faustman, W. O., and Mathalon, D. H. (2007) Synch Before You Speak: Auditory Hallucinations in Schizophrenia . *Am J Psychiatry*164:458–466.

Ford, J.M and Mathalon, D.H. (2004).Corollary discharge dysfunction in schizophrenia: can it explain auditory hallucinations? *Int J Psychophysiol*. 58(2-3):179-89

Frith C. (1987). The positive and negative symptoms of schizophrenia reflect impairments in the perception and initiation of action. *Psychological Medicine* 17:631-648.

Frith C. (1992). *The Cognitive Neuropsychology of Schizophrenia*. Hover: Lawrence Erlbaum Associates.

Frith, C (20121) Explaining delusions of control: The comparator model 20 years on. *Consciousness and Cognition*. 21: 52–54

Frith, C. 2005 The neural basis of hallucinations and delusions. *C. R. Biologies* 328: 169–175

Gallagher, S. (2004). Neurocognitive models of schizophrenia: A neurophenomenological critique. *Psychopathology* 37: 8-19

Haggard, P., & Clark, S. (2003). Intentional action: conscious experience and neural prediction. *Consciousness and Cognition, 12*(4), 695-707.

Hauser, M., Moore J. W., de Millas, W., Gallinat, J., Heinz A., Haggard, P., and Voss, M. (2011) Sense of agency is altered in patients with a putative psychotic prodrome. *Schizophrenia Research* 126: 20–27.

Heinks-Maldonado T. H.,. Mathalon D. H., Houde J. F., Gray M. Faustman, W. O., and Ford, J. M. (2007) Relationship of Imprecise Corollary Discharge in Schizophrenia to Auditory Hallucinations. *Arch Gen Psychiatry* 64:286-296.

Helmholtz, H. V. (1860). *Treatise on Physiological Optics*. New York, Dover.

Hoerl, C. (2001). On Thought Insertion. *Philosophy, Psychiatry, & Psychology* 8: 189-200.

Hurlburt, R. T. & Akhter, S. A. (2008). Unsymbolized thinking. *Consciousness and Cognition* 17:1364–74.

Hurlburt, R. T. (2009) Unsymbolized thinking, sensory awareness, and mindreading *Behavioral and Brain Sciences* 32: 29-30.

Jackendoff, R. (1996). How language helps us think. *Pragmatics and Cognition*, 4: 1-35.

Jackendoff, R. (2012) *A user's guide to thought and meaning*. New York: Oxford University Press.

Jackson J.H. (1958) *Selected Writings of John Hughlings Jackson*. New York: Basic Books.

Jackson, M.C., & Fulford, K.W.M. (1997) Spiritual experience and psychopathology. *Philosophy, Psychiatry, and Psychology*, 4: 41-65.

Jeannerod, M. (2003). The mechanism of self recognition in humans. *Behavioral and Brain Research* 142: 1–15.

Langdon, R., Jones, S. R., Connaughton, E. and Fernyhough, C. (2009) The phenomenology of inner speech: comparison of schizophrenia patients with

auditory verbal hallucinations and healthy controls. *Psychological Medicine*: 39, 655–663.

Langlad-Hassan, P. (2008) Fractured Phenomenologies: Thought Insertion, Inner Speech, and the Puzzle of Extraneity. *Mind and Language* 23: 369-401.

Lysaker J. , and Lysaker P (2005). Being interrupted : the self and schizophrenia. *Journal of Speculative Philosophy* 19:1–21.

Martinez-Manrique, F. & Vicente, A. (2010). What the...! The role of inner speech in conscious thought. Journal of Consciousness Studies, 17: 141-167.

Moritz, S. and Larøi, F. (2008) Differences and similarities in the sensory and cognitive signatures of voice-hearing, intrusions and thoughts. *Schizophrenia Research* 102: 96–107.

Pacherie, E. (2007) The anarchic hand syndrome and utilization behavior: a window onto agentive self-awareness. *Functional Neurology* 22: 211-217.

Pacherie, E. (2011) Self-Agency, in S. Gallagher, ed., *The Oxford Handbook of the Self*. Oxford University Press.

Peacocke, C. (2007). Mental Action and Self-Awareness (I), in J. Cohen and B. McLaughlin, eds., *Contemporary Debates in the Philosophy of Mind*, Blackwell.

Proust, J. (2006) Agency in Schizofrenia from a Control Theory Viewpoint, in N. Sebanz and W. Prinz, eds., *Disorders of Volition*, Bradford Books.

Sperry , R.W. (1950) Neural basis of the spontaneous optokinetic response produced by visual inversion . *Journal of Comparative and Physiological Psychology* , 43 , 482 –489 .

Stephens, G. and Graham, G. (2000). *When self-consciousness breaks: Alien voices and inserted thoughts*. Cambridge, MA: MIT Press.

Synofzik, M. and Vosgerau, G. (2008). Beyond the comparator model: A multifactorial two-step account of agency. *Consciousness and Cognition* 17: 219–239.

von Holst & Mittelstaedt (1950). Das Reafferenzprinzip . *Die Naturwissenschaften*, 20 , 464 – 476 .

Vicente, A. & Martinez-Manrique, F. (2011). Inner Speech: Nature and Functions. Philosophy Compass, 6: 209-219.

Vosgerau, G., and Newen, A. (2007). Thoughts, motor actions, and the self. *Mind and Language* 22: 22–43.