SUBJECTIVE MISIDENTIFICATION AND THOUGHT INSERTION

(Forthcoming in *Mind and Language*)

Matthew Parrott
King's College London

Non-Citable Draft

Abstract: This essay presents a new account of thought insertion. Prevailing views in both philosophy and cognitive science tend to characterize the experience of thought insertion as missing or lacking some element, such as a 'sense of agency', found in ordinary first-person awareness of one's own thoughts. By contrast, I propose that, rather than lacking something, experiences of thought insertion have an additional feature not present in ordinary conscious experiences of one's own thoughts. More specifically, I claim that the structure of an experience of thought insertion consists of two distinct elements: a state of ordinary first-person awareness and a sense that this state of awareness is highly unusual. In addition to modeling the experience of thought insertion, I also explain how a delusional pattern of thinking could lead someone who has this kind of experience to adopt a belief that some other entity is inserting thoughts into her mind. Finally, I briefly sketch a neurocomputational framework that could be developed to explain the sense that one's state of first-person awareness is highly irregular.

Thought insertion is a fairly common symptom of schizophrenia in which an individual claims to experience certain thoughts as inserted into her mind, most often by other people

Previous versions of this essay have been presented at Berkeley, King's College London, the University of Fribourg, the University of Manchester, and the University of Oxford. On all of these occasions, I was fortunate to receive extremely useful feedback and comments, for which I am very grateful. I would also like to thank Tim Bayne, John Campbell, Ellen Fridland, Anil Gomes, Nick Jones, Mike Martin, Eliot Michaelson, Ian Philipps, Hanna Pickard, Nick Shea, Josh Shepherd, Michael Sollberger, James Stazicker, Gottfried Vosgerau, and especially Martin Davies.

Address for Correspondence: Matthew Parrott, Department of Philosophy, King's College London, The Strand, London WC2R 2LS
Email: matthew.parrott@kcl.ac.uk

but sometimes by inanimate objects. To get a better sense of this, consider the following reports wherein subjects attempt to describe what is like for them to experience thought insertion:

I) Thoughts come into my head like 'Kill god'. It's just like my mind working, but it isn't. They come from this chap, Chris. They're his thoughts. (Frith, 1992)

II) I have never read nor heard them; they come unasked; I do not dare to think I am the source but I am happy to know of them without thinking them. They come at any moment like a gift and I do not dare to impart them as if they were my own. (Jaspers, 1963)

III) Sometimes it seemed to be her own thought '. . . but I don't get the feeling that it is.' She said her 'own thought might say the same thing . . . But the feeling isn't the same . . . the feeling is that it is somebody else's . . . (Hoerl, 2001)

IV) I didn't hear these words as literal sounds, as though the houses were talking and I were hearing them; instead, the words just came into my head – they were ideas I was having. Yet I instinctively knew they were not *my* ideas. They belonged to the houses, and the houses had put them in my head. (Saks, 2007)

In each of these passages, a subject is attempting to characterize his or her conscious experience, but the reports are extremely puzzling.[1] How could anyone really think that

[1] See Mullins and Spence, 2003; Sims, 2003; and Vosgerau and Voss, 2014 for similar reports. Campbell, 2001 and Jaspers, 1963 both suggest that reports like these are so bizarre that they may actually be meaningless. The difficulty one has taking them literally also motivates certain non-doxastic accounts of delusion (for example,

thoughts woven into their stream of consciousness belonged to someone else? What could someone possibly experience that would lead her to believe that some other entity has been inserting thoughts into her mind? Indeed, regardless of why it may have occurred to them in the first place, why would a person not immediately reject the idea that a foreign agent is inserting thoughts into her mind?

As with many delusions, there are at least two respects in which the phenomenon of thought insertion is extremely puzzling.[2] First, there is a phenomenological question about the character of the subject's experience: what is it like for someone to undergo an experience of thought insertion? But, in addition, there is a second question about the cognitive processing implicated in thought insertion: why does someone who undergoes an experience of thought insertion come to believe the sorts of things reported in the previous passages? Thus, even if we understand the structure of the experience of thought insertion (i.e., we understand what the experience is like), the phenomenon as a whole will remain obscure without some sense of why an agent undergoing such an experience adopts the

---

Currie, 2000). However, most theorists attempting to explain thought insertion try to take reports like these seriously. I believe the discussion in section four of this essay should help to make them seem more intelligible.

[2] This way of framing the issue fits well with the prevailing 'two-factor' approach to explaining delusions (cf. Coltheart, et. al. 2011; Coltheart, 2007; Davies, et. al. 2001). It is important to recognize, however, that I have deliberately left open how many cognitive 'factors' or impairments are implicated in thought insertion. I have claimed only that there are at least two aspects of the phenomenon that are likely to strike one as immediately puzzling. A third way in which thought insertion could seem puzzling is that it is not very clear why subjects maintain their belief that another entity is inserting thoughts into their mind in the face of counterevidence. Since this essay is concerned with the onset of thought insertion, it shall not address this third question.

belief that, for example, ideas in her own mind belong to houses on the street.[3] This is partly because we have a strong intuition that even an extremely bizarre experience would not justify, and so would not clearly explain, someone believing something which strikes us as impossible, or at least exceedingly improbable, namely that a foreign entity is inserting thoughts into her mind.[4]

In this essay, I shall present a model of thought insertion that aims to address both of these questions. I shall propose that the conscious experience implicated in thought insertion consists of two distinct elements: (i) a state of first-person awareness of one's thought that is of the same fundamental kind as ordinary conscious awareness of one's thoughts and (ii) a sense that this state of awareness is highly unusual or irregular. According to this proposal, the content of an experience of thought insertion is essentially first-personal: a subject is aware of *herself* (*de se*) thinking some particular thing. But she also experiences this first-person content as something highly unusual, which is what distinguishes the experience of thought insertion from an ordinary conscious experience of one's own thinking. This is a noticeable departure from most accounts in philosophy and cognitive science, which conceive of the experience of thought insertion as missing some feature found in ordinary conscious experiences of one's own thoughts.

---

[3] As we will see in the following section, some theorists argue that an experiential anomaly is the only cognitive deficit implicated in thought insertion and that subjects simply accept the contents of extremely bizarre experiences, in the same way they would accept the contents of ordinary experiences (cf. Sollberger, 2014). Such an 'endorsement' account promises a simple and straightforward way of addressing this second question about belief adoption, but, as we will see, it has far more difficulty with the first question.

[4] I shall assume that delusions are beliefs. For arguments supporting this assumption, see Bortolotti, 2010 and Bayne and Pacherie, 2005.

In section one I shall argue that standard approaches to explaining the experience of thought insertion neglect a kind of ambivalence typically expressed by subjects of the experience. In section two, I shall propose that the presence of this ambivalence suggests that, rather than missing something, an experience of thought insertion has some additional feature not present in an ordinary conscious experience of thinking. As we have just seen, however, having an unusual experience does not yet explain why a subject of thought insertion mistakenly attributes her thought to some other agent. Indeed, as we will see in section three, thought insertion has seemed especially perplexing to philosophers because it appears to contradict a highly regarded philosophical principle, which maintains that it is impossible for an individual to misidentify the subject of her own thoughts, at least when she has first-person awareness of them.[5] However, in section four, I shall draw an analogy with Sydney Shoemaker's conception of quasi-memory in order to illustrate a pattern of reasoning that I claim would naturally lead someone to believe some other agent is thinking her thoughts. The point of this analogy is not to illustrate any conscious reasoning undertaken by a subject of thought insertion but to present an intelligible model of the sub-personal cognitive transition from an unusual experience of one's thought to a delusional belief about a foreign agent inserting that thought. Finally, in section five, I shall briefly sketch a neurocomputational framework that can plausibly explain the experienced sense of abnormality, which I claim is present in experiences of thought insertion.

1. Contrasting Thought Insertion with Ordinary Experience

[5] The idea that thought insertion poses a threat to the immunity thesis was, as far as I know, first suggested by Campbell, 1999.

1.1 Ambivalence

A subject experiencing thought insertion is clearly alienated from his or her thoughts. But, in at least one important sense, she still experiences them as her own: she experiences them as located inside of *her* mind. She does not report being aware of another person's thoughts in virtue of something resembling what we might think of as a kind of telepathy. Nor does she experience the inserted thoughts as complete interruptions to her stream of consciousness. Instead, an inserted thought is experienced as related in some manner to her other thoughts and experiences (cf. Bayne, 2010). Normally thoughts located within a person's consciousness are hers, which is perhaps why these subjects tend to have a lingering sense of owning their inserted thoughts—the subject owns the mind in which the inserted thoughts occur.[6] However, in spite of this, subjects also quite clearly report that they are experiencing thoughts which do not belong to them.

Because they typically attempt to characterize their unusual experiences in both these ways, it seems that subjects experiencing thought insertion express a kind of ambivalent attitude. Recall the reports with which we began. Frith's subject explicitly describes experiencing his inserted thoughts as 'just like my mind working', but also not like it. The woman mentioned by Hoerl claims her thoughts seem to be her own but they also feel different; they feel as if they are someone else's. And, in Saks's autobiography, she describes

[6] Compare having another person's organ inserted into your body. In one sense, it seems to become yours in virtue of your owning the body it is within (perhaps also coupled with the donor's voluntary consent). However, the feelings surrounding an organ transplant are bound to be complicated. A friend has had one of his heart valves replaced and now feels at once that the new valve is his and, in another sense that it is not. This kind of ambivalence may be amplified in cases where receiving an organ causes noticeable physiological changes in the recipient (cf. Kuhn, et. al. 1988).

how thoughts come into her head just like 'ideas I was having'; however, she simultaneously understands 'instinctively' that the experience is not of her own thoughts, but of those belonging to houses. In each case, the person reporting what it is like to experience thought insertion expresses a kind of ambivalence, which suggests that something about the experience of thought insertion generates an ambivalent attitude. I do not wish to claim that such an attitude is essential to the experience but I do think it is sufficiently prevalent in first-hand accounts to offer valuable guidance as to the phenomenal character of the experience. It suggests that there are at two central elements to a conscious experience of thought insertion, one of which is familiar and the other highly unusual.

It may be worth noting how the type of ambivalence we find in cases of thought insertion differs from another sort of ambivalence an individual might have toward her own thoughts. It seems fairly common for people to occasionally have sudden unbidden or intrusive thoughts. However, in obsessive compulsive disorder (OCD), subjects experience intrusive thoughts in a much more serious way.[7] In OCD, intrusive thoughts are typically experienced as inappropriate or in conflict with a subject's values, beliefs, dispositions, or self-conception (Doron and Kyrios, 2005; Rowa, et. al. 2005; Rowa and Purdon, 2003). Thus, subjects of OCD feel somewhat alienated from obsessional intrusive thoughts because they strongly reject the contents of those thoughts. It is partly for this reason that obsessional intrusive thoughts generate strong feelings of anxiety or disgust, and subjects repeatedly try to suppress them. Nevertheless, subjects of OCD do not fail to recognize that unwanted obsessional thoughts belong to them (Abramowitz, et. al. 2009). They therefore

[7] It is unclear whether the intrusive thoughts which occur in obsessive compulsive disorder are on a spectrum with more typical intrusive thoughts or whether they are categorically different. For critical discussion of experimental work on this question, see Julien, et al. 2007.

seem to manifest a sort of ambivalence: they feel extremely alienated from intrusive thoughts which they nevertheless acknowledge to be their own.

Like subjects who experience obsessional intrusive thoughts, subjects of thought insertion also tend to disavow the contents of inserted thoughts, often judging explicitly that they are unwarranted. So could this be the source of their feelings of ambivalence or alienation? In recent years, a handful of philosophers have proposed to explain thought insertion on the basis of the fact that an individual experiences her thoughts as alien because she judges them to be rationally unwarranted.[8] Hanna Pickard (2010) offers the best defense of this line of thought:

I propose that schizophrenics disown mental events that seem to be manifestations of mental states that they do not, for some reason or other, endorse. Taking up the practical stance, looking outwards to the world, they judge that the mental states which these thoughts, impulses, or feelings bring to consciousness are not warranted or appropriate: they do not reflect how the world actually is or should be. (2010, pg. 67)

According to Pickard's proposal, conscious thoughts express a person's underlying mental states, which typically are states she judges to be rationally appropriate. In thought insertion, however, Pickard thinks that thoughts express mental states a subject judges, perhaps even implicitly, that she ought to not have. As Pickard puts it, 'identification occurs when a mental state is causally responsive to one's rational will. Alienation occurs when a mental state is not.' (2010, pg. 68)

[8] Bortolotti and Broome, 2009 and Fernandez, 2012 also attempt to explain thought insertion in terms of rational endorsement.

As should be clear, Pickard's account does not distinguish cases of thought insertion from cases of obsessional intrusive thought (or even from non-pathological cases of intrusive thought), in which subjects 'judge that the mental states which these thoughts, impulses, or feelings bring to consciousness are not warranted or appropriate.' So it does not seem that Pickard's framework can explain why an individual experiences thought insertion, rather than intrusive thoughts.

She herself takes this to be a virtue. Rather than sharply distinguishing thought insertion from other phenomena, Pickard believes that experiencing a thought to be in conflict with one's sense of reasons is a matter of degree. The phenomenon of thought insertion is, on her view, just an extreme case of someone becoming aware of a thought she does not rationally endorse. By recognizing similarities between thought insertion and other, more ordinary, experiences, Pickard thinks we can better explain how we are capable of empathizing with subjects of thought insertion, which she claims is crucial to effectively treating them.

However, by assimilating thought insertion to cases of intrusive thought, Pickard overlooks an important difference between the sources of alienation and ambivalence in the two cases. As we can see in OCD, a person who judges *P* to be unwarranted or inappropriate tends to feel disconnected or alienated from *P* and she may therefore feel a degree of ambivalence toward any intrusive thought representing *P*. But notice that in this case her feeling is grounded in the thought's content. Her rational disavowal of *P* is what explains her feeling ambivalent toward particular thoughts representing *P*. In other words, the subject is estranged from her thoughts only because she feels alienated from their contents--the former seems to depend on the latter. With respect to thought insertion, however, the converse seems true. Any feeling of alienation that one may have with respect

to the content of an inserted thought seems to depend on experiencing the *bearer* of that content as alien or inserted, regardless of whether the content is taken to be rationally warranted. Subjects of thought insertion report feeling alienated from mental particulars, sometimes to the point of reporting that they feel them being pushed into their head. From their point of view, the rationality of the content of the inserted thought seems secondary.[9]

It therefore seems that there is a significant phenomenological difference between experiences of inserted thoughts and experiences of obsessional intrusive thoughts. Any ambivalence expressed by a subject of thought insertion seems directed primarily at a particular thought token, a specific act, occurrence, or episode of thinking, rather than at the content of a thought. So if there is an ambivalent attitude generated by the experience of thought insertion, it too is primarily about a particular episode of thinking, rather than about the content represented by that episode. It is the presence of this type of ambivalent attitude that I think is characteristic of an experience of thought insertion and it suggests that the phenomenal character of the experience of an episode of thinking is structured in a way that is both familiar in one respect but a highly unusual in another.

To get a more nuanced picture of the phenomenology of the experience, it will be useful to contrast an experience of thought insertion with that of an ordinary conscious experience of thinking. Many philosophers and cognitive scientists pursue this strategy and they tend to characterize the former as missing some central feature of the latter.[10] In the

[9] Another way to see this point is to consider what would happen if a subject of thought insertion were to conclude that the content of an inserted thought was actually fully justified. It would, I think, be surprising if she then ceased thinking of the thought as inserted.

[10] Despite notable differences, a number of theories maintain that the phenomenal character of an experience of thought insertion is essentially missing something found in an ordinary experience of one's own thought.

following subsection, I shall focus on one example of this approach, an account of thought insertion presented by Christopher Peacocke (2008). I shall argue that Peacocke's account is incomplete precisely because it conceives of thought insertion in negative terms--as lacking features of an ordinary experience of one's own thought. Although I lack sufficient space to discuss similar proposals which characterize the experience exclusively in terms of what it is missing, it should be clear that they too will face the problem of being incomplete.

1.2 Peacocke on Conscious Experience of Thought

Philosophers have long thought that we have a special way being aware of our own thoughts.[11] Indeed, it seems natural to think that part of what it is to have a conscious thought is to be aware of it in a distinctively first-personal way. More recently, some philosophers have sought to develop this idea by emphasizing that many of our thoughts are actually mental actions (e.g., O'Shaughnessy, 2000; Soteriou, 2013). For example, we engage in actions by judging, doubting, intending, imagining and forming beliefs; even entertaining a thought could be construed as a type of mental action (cf. Roessler, 2013). These share with bodily actions the fact that we have a special first-person way of being aware of them - put

---

Most commonly, several theorists characterize thought insertion in terms of a missing 'sense of agency', including Coliva, 2002; Frith, 1992; Gallagher, 2004, 2007; Pedrini, forthcoming; Stephens and Graham, 2000; Sousa and Swiney, 2013; and Vosgerau and Voss, 2014.

[11] Sometimes this first-person mode of awareness is identified as 'introspection'. I am intentionally avoiding this term because of its associations with certain substantive theories of self-knowledge. Specifically, it might connote that one is aware of one's thought in virtue of having a higher-order thought or perception of it. I want to avoid that connotation. In this essay, I use the term "first-person awareness" in the most neutral way possible. I mean only to refer to the distinctive way each of us is normally consciously aware of our thoughts, whatever that may be.

most simply, we are aware of them solely in virtue of being their agent. Normally, if I judge that it is unseasonably cold or imagine myself sitting by a warm fire, I am immediately aware that I am so judging or imagining.

According to Peacocke, it is because many of our thoughts are mental actions that we have a special way of being aware of them, which he calls 'action-awareness'. There are three things about Peacocke's conception of action-awareness that are most relevant to our discussion. First, Peacocke thinks that action-awareness is the fundamental way in which a person ordinarily comes to know about her own acts of thinking, about those of her thoughts that she actively produces. However, a state of action-awareness is not itself an epistemic state – it is not knowledge or belief that one is thinking something. It is rather the basis on which one normally judges that one is thinking something. Second, the content of a state of action-awareness is essentially first-personal or *de se*: it is <I am thinking θ now>[12]. This is to say that my conscious experience has the content of *myself* thinking a particular thing.[13] Because the content is essentially *de se,* if one simply endorses that content, or takes it at face value, she will thereby know that *she* is thinking such and such now. As Peacocke puts it, 'you are entitled to take for granted the existence of the objects of their singular intentional contents, including their *de se* content.' (2014, pg. 127) Finally, the *de se* element of a subject's conscious experience is plausibly non-conceptual; it is what Peacocke calls a

[12] θ is a variable for a particular thought. For those that think thoughts always involve propositional contents, one can substitute 'that *P*'.

[13] Again, this should not be confused with a sort of higher-order theory on which there are two states, for instance my thought that it is raining and my thought that I am thinking that it is raining. On Peacocke's view, the first-order state has *de se* content.

notion rather than a concept. One therefore does not need to acquire a specific level of conceptual sophistication in order to be conscious of *herself* thinking θ now.

Peacocke proposes that experiences of thought insertion completely lack this first-person action-awareness:

What the schizophrenic subject lacks in the area of conscious thought is action-awareness of the thoughts that occur to him. To enjoy action-awareness of a particular event of thinking is to be aware, non-perceptually, of that thinking as something one is doing oneself. The awareness of one's own agency that exists in normal subjects is missing, in, for example, the schizophrenic experience of 'thought-insertion.' (2008, pg. 276)

The suggestion in this passage about a lack of 'awareness of one's own agency' is noticeably similar to proposals that attempt to explain thought insertion in terms of a missing 'sense of agency'. Compare, for instance, Shaun Gallagher's claim that 'whatever the precise nature of the neurological disruptions for delusions of control or thought insertion, some such processes clearly generate a first-order phenomenal experience that lacks a sense of agency.' (2007, pg. 42) Although it is not uncommon to hear thought insertion described in these terms, notice how claims about a missing 'sense of agency' are potentially ambiguous. A 'sense of agency' could be construed as some kind of phenomenal or sensory experience of one's own agency, which is how Gallagher intends it. However, it could also be thought of as a kind of awareness of oneself as the agent or producer of a particular thought. Peacocke's approach naturally suggests the latter interpretation; it suggests that thought insertion results from a disturbance or interruption to a subject's awareness of herself as the agent of her own thinking.

There is an important assumption in the background of Peacocke's account that is worth making explicit. The claim that a subject who experiences thought insertion lacks 'the awareness of one's own agency that exists in normal subjects' would not be explanatory unless the thought in question is a mental action. Yet, as Peacocke seems to acknowledge in recent work, not every conscious thought is most plausibly construed as an action.[14] In addition to making judgments and forming beliefs, we naturally classify passive mental phenomena such as tunes that run through one's head, or sudden surprising ideas as thoughts. Since these are not mental actions, they too would occur without 'the awareness of one's own agency that exists in normal subjects'.[15] For this reason, one might worry about the explanatory power of Peacocke's proposal. Merely postulating an absence of action-awareness fails to distinguish cases of thought insertion from passive episodes of thinking, which obviously do not generate delusions.[16] However, I think the correct lesson to draw from these observations is that an absence of action-awareness would not be a manifestation of any cognitive impairment except in cases where one's thought is a mental action. Once we distinguish episodes of thinking that are actions from those that are not, it is clearer how an absence of a certain mode of awareness in the one case may be highly irregular, while in the other case it would be perfectly ordinary.

---

[14] For instance, he claims that a thinker can engage in *cogito* reasoning without experiencing himself 'as the agent of the thinking which forms its starting point' (2014, pg. 129). For doubts about this, see Roessler, 2013.

[15] This is not to say that Peacocke has no means of explaining conscious awareness of passive episodes of thinking (see Peacocke, 1998).

[16] Bortolotti and Broome, 2009 and Vosgerau and Voss (2014) both raise this sort of objection to accounts that describe thought insertion in terms of a missing 'sense of agency'.

Nonetheless, as it stands, Peacocke's account of thought insertion is incomplete. This is because he characterizes the phenomenal character of the experience in purely negative terms. In saying that a subject's experience of an inserted thought lacks action-awareness, he has not given us any clue as to what the experience is like. How is a subject aware of her inserted thought if not via first-person action-awareness? Does she have some other kind of conscious access to the thought? Sometimes it seems that Peacocke wants to appeal to a kind of analogy by drawing attention to the fact that subjects who experience thought insertion also lack action-awareness of bodily-actions. But, since it is quite obvious that one can be perceptually aware of her bodily actions, it is easy to understand how a person could become aware of a bodily action while lacking action-awareness of it. By comparison, it is implausible to think that subjects of thought insertion are perceptually aware of inserted thoughts; so the analogy with bodily actions does not really illuminate what these experiences are like. Perhaps lacking action-awareness of one's thoughts just means that the content of one's experience is missing a *de se* element, but then it really would be difficult to distinguish experiences of thought insertion from passive episodes of thinking. We therefore need a more positive description of the phenomenal character of an experience of thought insertion to supplement the negative characterization Peacocke offers.[17]

[17] Peacocke recognizes this and claims that his account will need to be supplemented with some kind of empirical explanation. In his estimation, 'a full understanding has to explain the prevalence of the impression of control by alien agencies and forces. Why an absence of action-awareness should lead to this specific kind of illusion needs an empirical explanation…' (2008, pg. 279) Peacocke himself thinks the right approach will appeal to similarities between experiences of thought insertion and auditory verbal hallucinations. However, although many schizophrenic subjects do experience auditory verbal hallucinations, these seem to be quite different from their experiences of inserted thoughts. It is at least true that individuals take the two to be very

1.3 The Endorsement Account

One option for developing such a view would be to pursue the idea that an experience of thought insertion literally represents a thought being inserted into one's mind by some external agent. This would be a kind of 'endorsement account' of thought insertion along the lines sometimes offered to explain other monothematic delusions. On this sort of view, the content of an experience of thought insertion would literally represent <so and so inserting θ into my mind> or perhaps <so and so thinking θ>. This would thereby fully account for the pathology of thought insertion at the experiential level.[18] Furthermore, adopting an endorsement account would have the advantage of helping us understand why a subject attributes her inserted thoughts to a *specific* external agent, to, for instance, 'Chris' or to a particular house. The specificity at the level of judgment would be explained by the fact that the subject is simply taking the rich content of her experience at face value.

Although an endorsement account may be plausible for some delusions, I think it oversimplifies what an experience of thought insertion is like.[19] Recall the individual who reported that 'it's just like my mind working, but it isn't'. Let's suppose this experience does have the content <Chris is inserting thoughts into my head>. This might explain some

---

different kinds of experience and there does not seem to be any good reason to doubt this aspect of their reports (cf. Nayani and David, 1996).

[18] For a defense of an endorsement account of thought insertion, see Sollberger, 2014.

[19] Another difficulty for the endorsement account is that it predicts a wider range of unusual experiences than we actually find (Davies, et. al, 2001). If there were some neural impairment causing these sorts of experiences, we would expect more than just the experience of <*a* is inserting θ into my head>. For further criticism of endorsement accounts generally see Coltheart, 2005, for a general defense of the approach, see Bayne and Pacherie, 2004.

aspects of his report but it would not explain why he takes the experience to be 'just like' his mind working. Indeed, it would be very unclear why he felt any ambivalence toward the inserted thought at all. Why would he not straightforwardly report the content of his experience or, at least, report that that is how things seemed? Perhaps a defender of the endorsement account could try to capture the subject's ambivalence, but it is very hard to see how this sort of thing could be represented in the content of one's experience. How are we to make sense of an experience explicitly representing a thought as being inserted by another and also as being 'just like my mind working'? How could these both be represented within the content of a single experience? By supposing that every aspect of what a subject reports is somehow represented in the content of her experience, the endorsement account risks obscuring our understanding of the phenomenon.

In response to an earlier version of this criticism, Michael Sollberger has recently argued that there are two ways for an endorsement theorist to make sense of the ambivalence expressed by subjects of thought insertion. First, he thinks it could be explained by 'reference to other features that delusions of thought insertion possess,' specifically by 'the patient's experiencing a dissociation' between a sense of ownership and a sense of agency. (2014, pg. 607) But it isn't clear to me why a subject of thought insertion would experience this sort of dissociation if, as the endorsement theorist claims, the content of her experience is something like <so and so inserting θ into my mind>. Since the content of that experience explicitly represents a foreign agent producing θ, any absence of a sense of agency would not be unusual. But neither would the presence of a sense of ownership. After all, the content of the experience represents θ being inserted into *my mind*. So why would I not have some sense that I own it? It therefore seems to me that these 'other

features' of the delusion don't really illuminate how the experiential content posited by the endorsement theorist would generate feelings of ambivalence.

The second strategy Sollberger suggests seems even less promising. This is for the endorsement theorist to simply insist that the ambivalence found in thought insertion 'can be grounded in the experiential content' because, like certain visual illusions, 'delusional experiences of inserted thoughts…involve contradictory contents' (2014, pg. 608). I have some doubts about whether perceptual experiences can represent explicit contradictions. But, regardless, what is the contradictory content supposed to be in experiences of thought insertion? The endorsement account's claim is that the content of the experience is something like <so and so inserting θ into my mind>. That is not a contradiction. Maybe Sollberger's idea is that the experience also represents the subject producing θ herself; i.e., it also has the content <I am producing θ >.[20] But if so, then the endorsement theorist faces another problem. Since subjects of thought insertion clearly do not endorse both of these contents, the endorsement theorist owes some explanation as to why they endorse only one of the two contradictory contents represented by their experience. It is not at all clear to me what such an explanation would look like. So it seems to me that this second strategy does not offer much help in terms of capturing the ambivalence expressed by subjects of thought insertion.

There may be some other way to supplement Peacocke's negative description of the phenomenal character of an experience thought insertion.[21] But, rather than exploring other

[20] It is not enough to think the contradictory content is <I am thinking θ> if this means that θ is in my mind, for that is already part of the content and <so and so inserting θ into my mind> and so not a contradiction.

[21] Indeed, Peacocke's account could be supplemented by the sort of neurocomputational framework I sketch in section five. Therefore, the key difference between Peacocke's account and the one presented in this essay is

alternatives, in the following section I would like to suggest that we understand the phenomenology of thought insertion in a different way.

2. The Structure of an Experience of Thought Insertion

We have seen that individuals experiencing thought insertion tend to express ambivalence toward inserted thoughts. But what might someone be trying to convey by expressing this ambivalence? How *could* an experience be 'just like' my mind working and also, at the same time, not like my mind working? I would like to suggest that a person experiencing thought insertion has ordinary first-person awareness of a particular act of thinking. In keeping with Peacocke's agentive model, this means she has a conscious state with *de se* content <I am thinking θ now>. Thus, the content of an experience of thought insertion is of the same fundamental kind as that of an ordinary conscious experience. But I also would like to propose that the phenomenal character of an experience of thought insertion has some additional structure. Specifically, I propose that the following two claims characterize a conscious experience of thought insertion for an individual, *a*:

(A) *a* has first-person awareness of <I am thinking θ now>.

(B) *a* has a sense that her state of awareness in (A) is *not* ordinary first-person awareness.[22]

---

whether or not an experience of thought insertion is of the same fundamental kind as an ordinary experience of one's thought. On Peacocke's view it is not, whereas, according to the view presented in this essay, it is.

[22] We may be able to give a more determinate gloss to this; however, at this stage it is best to put (B) in general terms. For further discussion, see Ellis and Young, 1990 and, especially, Maher, 1999.

According to this model, we should not think of experiences of thought insertion as missing something normally found in conscious experiences of one's thoughts, but as having some kind of extra phenomenal overlay that an ordinary conscious experience of thinking lacks. The sense or feeling in (B) does not typically accompany awareness of one's own thoughts. It therefore signals something extraordinary, something radically different from ordinary awareness. Moreover, it is because of (B) that *a* does not take the content of her experience in (A) at face value. She doubts that she is thinking θ even though that is *exactly* how things are represented in her state of awareness. In this way, (B) functions as a kind of misleading counter-evidence to the content in (A).

This proposal can help us make sense of the ambivalence expressed by subjects who experience thought insertion. (A) and (B) schematically capture each side of that attitude. Having first-person awareness of <I am thinking θ now> *is* just like one's own mind working; it is exactly like it. We have the same kind of awareness in non-pathological cases. But, having the sense in (B) makes the total phenomenal character of the experience very different from ordinary thinking. The sense in (B) therefore alienates the person from her state of awareness and this helps explain why her experience does not feel 'the same' as the ordinary case. What the total experience of thought insertion is like for an individual therefore involves more than what is represented in the content of her experience. The sense in (B) contributes to the phenomenal character without figuring directly in its content. In this way, the entire experience is at once familiar and foreign; the thinking seems to be one's own but also seems to be not one's own. Based on their reports, this seems to be what subjects of thought insertion are trying to tell us their experiences are like.

3. Subjective Misidentification and Immunity to Error

According to the proposal presented in previous section, the content of (A) involves a *de se* representation, so it follows that anyone experiencing such a content does not need to make an identification judgment in order to know *who* is thinking θ. As Peacocke stresses, 'when you take such an awareness at face value…your judgment is identification-free.' (2008, pg. 248) So, if a subject of thought insertion has first-person awareness of <I am thinking θ now> but nevertheless goes on to misidentify the subject of her thought, she must not be taking that experience at face value. But even if she were alienated from her experience because of the sense in (B), why would she attribute θ to some other entity?

Many philosophers have been especially intrigued by this last question because thought insertion seems to be a straightforward counterexample to the principle that attributions of thought, when made on the basis of first-person awareness, are immune to errors of misidentification. This is, for example, what John Campbell is alluding to when he remarks that 'a patient who supposes that thoughts have been inserted into his mind by someone else is right about which thoughts they are, but wrong about whose thoughts they are.' (1999, pg. 620; cf. Coliva, 2002) Campbell is clearly right that the subject of thought insertion misidentifies the subject of a particular thought: she wrongly judges that the subject of one of her thoughts is some other agent. Yet, she also seems to know the content of the inserted thought. As Campbell claims, this suggests that the way in which we normally self-ascribe thoughts is not really immune to mistakes of misidentification.

To be clear, Campbell's worry is not merely that thought insertion would show that it is possible for an individual to misidentify the subject of her own thoughts. It should already be obvious that this can happen. We know that ascriptions of mental states made on

a third-person basis are open to mistakes of identification. For example, if I am at a party I might hear someone in the other room say that she thinks the party is boring. On the basis of this evidence, I might come to believe that Susan believes that the party is boring (perhaps I saw Susan enter the room a few moments earlier). Because I directly heard some person say what she thinks about the party, I at least know that *someone* believes the party is boring. Nevertheless, I might be wrong about that person being Susan. This is a paradigm of misidentification but it is also something that is clearly possible with respect to one's own thoughts whenever those are known about in a third-personal way. For example, if I hear a muffled recording of myself expressing the belief that London is a nice place to live, I could, on that basis, come to know that *someone* believes that London is a nice place to live but mistakenly think it is someone else, perhaps because I failed to recognize my own voice (cf. Ismael, 2012). Because self-ascriptions made on the basis of third-personal evidence are susceptible to these kinds of errors, it is only self-ascriptions of thought made on the basis of first-person awareness that appear to be immune to these sorts of errors of misidentification (cf. Coliva, 2006; Peacocke, 2014; and Recanati, 2012).

These considerations suggest the following formulation of the Immunity Thesis:

Immunity Thesis: If, on the basis of first-person awareness, a subject knows that someone is thinking θ, she cannot, on the same basis, be mistaken about who is thinking θ.

So formulated, it is not difficult to see why we might think cases of thought insertion pose a counterexample. If we assume that a subject of thought insertion does have ordinary first-person awareness of her thoughts, as the account presented in this essay holds, then it might naturally seem like she knows on the basis of this awareness that someone is thinking θ but

is mistaken about who is thinking it. If that is right, however, then the Immunity Thesis is mistaken.

Several philosophers, including Campbell, have tried to sidestep this threat of a counterexample by making a distinction between two different senses in which a subject can 'own' a thought (Campbell, 1999; cf. Stephens and Graham, 2000).[23] Indeed, it is now fairly standard to claim that someone can own a thought in the sense of being its agent or producer, but can also be the owner in a non-agentive or predicative sense.[24] The prevailing assumption seems to be that as long as we have two senses of thought ownership, we can make sense of the reports of thought insertion as involving a disruption in the agentive sense without a disruption of the other sense (e.g., the thought is still 'mine' in the predicative sense). This line of reasoning offers a way to save the Immunity Thesis from the threat of counterexample because we can maintain that although a subject of thought insertion knows that someone is thinking θ in a non-agentive or predicative sense, her misidentification involves the agentive sense. Thus, thought insertion would only appear to be a counterexample because of an equivocation between these two different senses.[25]

---

[23] For some additional doubts about the usefulness of this distinction, see Roessler, 2013.

[24] One might be inclined to think of the 'non-agentive' sense in terms of a thought's 'location' (cf. Bortolotti and Broome, 2009), but there is no reason to think that predicating a thought amounts to locating it in space.

[25]An exception to this strategy is Coliva (2002) who argues against Campbell's suggestion that there are distinct senses of thought ownership. Coliva thinks the distinction is not necessary because she insists that 'if someone makes a psychological self-ascription on the basis of introspection, then her judgment is logically IEM [immune to errors of misidentification]' (2002, pg. 33) This strong claim is motivated by Coliva's view that (i) introspection does not involve what she calls an 'identification component' and (ii) 'a judgment is liable to EM [errors of misidentification] if and only if its grounds contain an identification component.' (2002, pg. 28) As we will see in the following section, (ii) is mistaken. Another worry is that Coliva proposes that 'the reason why

It is not implausible to think there are multiple senses of thought ownership and distinguishing them does show how there is some logical space to defend the Immunity Thesis. The problem with this strategy, however, is that merely making this distinction does not really address the central question of why a subject of thought insertion actually does misidentify the subject of her own thought. The puzzle presented by thought insertion is that a subject attributes one of her own thoughts to some other entity in the face of a conscious experience of *herself* thinking θ now.[26] It is very difficult to see what reason someone could have for making this sort of mistaken attribution. The difficulty is not mitigated by reflecting on the possibility that the misidentification may involve different senses of ownership.

In the following section, I shall suggest a pattern of reasoning that plausibly characterizes the cognitive transition from an experience with the structure of (A) and (B) to the belief that one's own thought belongs to some other entity. As we will see, if a subject of thought insertion does respond to her anomalous experience in the way I suggest, one interesting consequence is that the phenomenon of thought insertion is actually compatible with the Immunity Thesis.

---

one cannot make an error through misidentification when one is self-ascribing a mental property on the basis of one's introspective awareness' is because 'the self-ascription is based on one's *being* in that very mental state.' (2002, pg. 28) It seems to me, however, that, rather than explaining it, this actually trivializes the Immunity Thesis: the reason one cannot make a mistake through misidentification is that one cannot make *any* kind of mistake when introspectively self-ascribing a mental state. For further discussion, see Campbell, 2002.

[26]It may be worth keeping in mind that a different way to avoid concerns about the Immunity Thesis would be to adopt an account like Peacocke's, which denies that experiences of thought insertion involve ordinary first-person awareness. If an experience of thought insertion were not partially constituted by first-person awareness, then the antecedent of the Immunity Thesis would not be satisfied (cf. Vosgerau and Voss, 2014).

4. Inserted Thoughts and Quasi-Memories

4.1 Odd Experiences and Delusional Beliefs

Even if we suppose that a subject with thought insertion has a highly anomalous experience with the structure outlined in section two of this essay, the adoption of a delusional belief that one's own thought has been inserted by some foreign entity can still seem unintelligible. This is a common reaction to an influential approach in cognitive neuropsychology that holds a subject adopts a delusion in order to explain the occurrence of a strange experience (Ellis and Young, 1990; Davies and Egan, 2013; Maher, 1974; Stone and Young, 1997). One standard criticism of this approach, as Pacherie and colleagues note, is that 'delusional beliefs are typically very poor explanations of the events that they are supposedly intended to explain.' (2006, pg. 567; cf. Davies, et. al. 2001; Campbell, 2001; Fine, et. al. 2007) The problem is that since we think the content of a subject's delusional belief is exceptionally improbable, it is far from clear how (or even that) it does explain the occurrence of an unusual experience. For example, given its overwhelming implausibility, why would someone take the belief that Chris is inserting thoughts into her mind to be a credible explanation of an experience of the sort described in section two? Why would she not adopt a more plausible explanation instead? The extremely low likelihood of someone inserting thoughts into one's mind suggests that there is some additional cognitive deficit or bias implicated in thought insertion (cf. Fletcher and Frith, 2009; Synofzik, et. al. 2008).

For this reason, theoretical models which focus exclusively on the phenomenological question of what it is like for someone to experience thought insertion cannot fully elucidate why an individual comes to believe the kind of things that subjects of thought insertion

report.[27] In what remains of this section, I shall attempt to shed some light on the cognitive transition from an experience with the structure of (A) and (B) to a delusional belief about another agent inserting thoughts. Although the pattern of reasoning I shall describe is delusional and in some sense not rational, I do not think it is unintelligible. Indeed, I think we can understand how someone may be susceptible to this form of reasoning because a well-known philosophical thought experiment exemplifies a rather similar pattern.[28]

4.2 Quasi-memories

Many years ago, Sydney Shoemaker introduced the concept of a quasi-memory. By definition, quasi-memories are experiences that are indistinguishable from genuine memories; which means that from a subject's point of view there is no phenomenal difference between quasi-remembering something and remembering it.[29] For the purposes of

---

[27] The exception, of course, is the aforementioned endorsement account. Since I have already registered difficulties for that account, I shall set it aside for the remainder of this essay.

[28] It is important to recognize that in what follows this reasoning is not conscious or occurring at the 'personal' level.

[29] Might this be because they are fundamentally the same kind of experience? One might think that the possibility of quasi-memory shows that the content of a genuine memory is thin. Rather than representing *me,* the apparent *de se* element of the representation is only a representation of something like 'as if' me. In this case, the content of a quasi-memory and an actual memory would be identical, which is why they would be indistinguishable. This is how Shoemaker conceives of quasi-memory (1970). Alternatively, one could adopt a kind of disjunctivism, maintaining that the content of memory and quasi-memory are fundamentally different even though they are subjectively indistinguishable. This is how Evans (1982) thinks of quasi-memory. In this essay, I shall assume that authentic memories are fundamentally different from quasi-memories because this preserves the intuition that memories are *essentially* a way of knowing about one's own past experiences. If the

this essay, let us also assume that memories have first-person contents *essentially*: they are *de se* representations or representations of *one's self* having done something in the past. Quasi-memories would have indistinguishable contents but, unlike genuine memories, they would be causally connected to the past of some other person. So, for example, if I genuinely remember watching the sunrise this morning, it follows that I did watch the sunrise this morning. But, if I merely quasi-remember watching the sunrise this morning, it would be someone else who saw it. If the right causal connections were in place, for example, I could quasi-remember that you watched the sunrise. It is just that this experience would be subjectively indistinguishable from my actually remembering my watching the sunrise.

Shoemaker brings up this notion of quasi-memory in part to make a point about conditions for self-identification. He writes:

Suppose that it were possible to quasi-remember experience other that one's own. If this were so, one might remember a past experience but not know whether one was remembering it or only quasi-remembering it. Here, it seems, it would be perfectly appropriate to employ a criterion of identity to determine whether the quasi-remembered experience was one's own. (1970, pg. 254)

According to the model we are working with in this essay, a genuine memory has *de se* content, <I experienced such and such then>. So if I really did remember watching the sunrise, I would only need to take my experience at face value in order to know that I did in fact watch the sunrise. However, in this passage Shoemaker suggests that if it were possible to quasi-remember things, then, even in cases where I did *actually* remember a past

---

two were not fundamentally distinct, it would be reasonable to think memory could give us knowledge only of an existential proposition, not a *de se* proposition.

experience, I would not be able to discern whether I was 'remembering it or only quasi-remembering it.' In such a case, it would be reasonable for me to withhold judgment about my memorial experience until I ruled out its being a mere quasi-memory, which, as Shoemaker notes, would require me to 'employ a criterion of identity'. Notice, however, that once I stop taking my memorial experiences at face value, my judgments become vulnerable to errors of misidentification. I open myself up to the possibility of taking a genuine memorial experience to be a *mere* quasi-memory, even though it is not.

The coherence of this conception of quasi-remembering depends on two things. First, it relies on a gap between the point of view of the subject currently having a memorial experience and the subject represented in the content of that experience. Even if the subject represented in my genuine memories is actually *me*, even if we stipulate that memorial experiences have *de se* contents *essentially*, there is room for me to question this aspect of my experience. This is true even if we assume that remembering watching the sunrise entails knowing that I watched the sunrise. Shoemaker's case would then illustrate that remembering (knowing) something does not guarantee that one knows *that* one remembers (knows).

Secondly, as Shoemaker points out, this sort of doubt about the *de se* component of my memory is only reasonable if quasi-remembering is a nearby possibility. Since there is in fact no nearby world in which I quasi-remember the past of others, it would actually be irrational of me to not take my memorial experiences at face value and believe precisely what they represent. However, if quasi-memory were a nearby possibility, then it would be reasonable to doubt whether or not *I* watched the sunrise solely on the basis of my memorial experience. Indeed, as Shoemaker says, it would be reasonable to 'employ a criterion of identity to determine whether the quasi-remembered experience was one's own.' The

essentially first-personal content of a particular kind experience therefore does not suffice to rule out the possibility of errors of misidentification in cases where one has good reasons for doubting that aspect of the content. This is why one cannot explain immunity to error through misidentification simply by claiming conscious thoughts and experiences have *de se* contents or, as Coliva (2002) claims, that first-person access does not require one to make any kind of identity judgment.

If a subject thinks quasi-remembering is possible, regardless of whether or not she is correct in this regard, it will rationally seem to her that she cannot know merely from the content of her memorial experience that she is remembering rather than quasi-remembering. She will therefore look for further justification or evidence to determine the identity of the person represented in the content of her experience. Crucially, when she does this, her judgment about who watched the sunrise will not be based solely on the memorial experience. It is because her judgment about who watched the sunrise is based on more than the *de se* content of her memorial experience that it is vulnerable to misidentification.

4.3 Believing in Inserted Thoughts and Immunity to Error

By analogy, if a person thinks it is possible for her to have direct awareness of another person's thoughts, if she thinks something like thought insertion is a close possibility, she will naturally think that she cannot know merely from the content of her experience that she is the one who is thinking θ now. This is true even though the content of her experience (A) is essentially first-personal. She would therefore refrain from taking the content of her experience at face value and, as in the quasi-memory case, look to determine the identity of the agent represented in the content of her experience. If she does this, however, her judgment about who is thinking θ will be based on more than her state of first-person

awareness. Like the case of quasi-memory, she will employ an additional 'criterion of identity' to determine who is thinking θ, which makes her vulnerable to errors of misidentification. If this is right, then, contrary to what some philosophers have feared, a case of thought insertion does not satisfy the antecedent of the Immunity Thesis. This is because the subject of thought insertion does not base her judgment about her putatively inserted thoughts solely on her state of first-person awareness, but rather appeals to additional evidence or criteria to determine who is thinking θ.

It might be objected that even if a subject experiencing thought insertion looks beyond her state of first-person awareness to determine precisely who is thinking θ, she would still know that *someone* is thinking θ solely on the basis of that experience, and this alone would be sufficient for the phenomenon to be a counterexample to the Immunity Thesis. Indeed, this looks like the proper conclusion to draw if we take the analogy with Shoemaker's case seriously. Someone who doubts the *de se* content of her genuine memorial experiences seems to nevertheless know that *someone* watched the sunrise solely on the basis of that experience. By definition quasi-memory is causally connected to someone's past, so having an experience that is either a quasi-memory or an actual memory of watching the sunrise looks like it is sufficient for knowing that someone watched the sunrise.

This objection brings out a crucial difference between the cases of thought insertion and quasi-memory. In Shoemaker's imagined case, it may be true that one is able to know the existential proposition that *someone* watched the sunrise on the basis of a genuine memory even if one doubts that the experience represents oneself.[30] But this is only true in cases

[30] Evans denies this, claiming that '*these judgments could not possibly constitute knowledge*.' (1982, pg. 244) But this denial assumes that the subject in question does not know that quasi-remembering someone else's past is a

where quasi-memories are nearby possibilities. Because quasi-memories are indistinguishable from memories, if they were epistemically possible, it would be plausible to think that having a memorial experience would enable one to know that her experience is either a memory or a quasi-memory. Crucially, however, the analogue of this epistemic situation does not hold for a subject experiencing thought insertion. There is no nearby world in which other people insert thoughts into peoples' minds and neither are experiences of thought insertion explicitly defined in relation to some other person's thinking. So there is no good reason for someone to doubt the first-person constituent of the experience <I am thinking θ now> that is not also a reason to doubt the existential content <*someone* is thinking θ now>. Once a person doubts that her experience accurately represents herself thinking θ now, she has no independent reason for thinking it succeeds in accurately representing some other person thinking θ. She is therefore not warranted in thinking that *someone* is thinking θ solely on the basis of her first-person awareness. It may be right to say that in some sense she is in a position to know this, but only if she first accepts the content of her first-person awareness and then correctly infers the existential claim from it. Once she denies the former, she no longer knows the latter.

The pattern of thinking I am describing in this section illustrates an additional way in which a subject with thought insertion manifests anomalous cognition (cf. Parrott, forthcoming) According to the transition I have described, a subject reasons from an experience with the structure of (A) and (B) to a misattribution of θ to some other agent because she (tacitly) accepts that thought insertion is a nearby possibility. That is to say a

---

nearby possibility. On that assumption, he may be right. But it may nevertheless be possible for a subject's to know an existential claim about the past if she were to *know* that quasi-remembering were a nearby possibility.

subject of thought insertion must think it is possible for other agents to be the source of contents like <I am thinking θ now> in order for this to be even a *potential* explanation of her unusual experience. So it is partly because a subject has an irregular conception of candidate explanations for her unusual experiences (a conception which includes the delusional one as a rather serious candidate) that she comes to believe that some foreign entity is inserting thoughts into her mind. [31]

Notice, however, that once we postulate that a subject of thought insertion has an irregular conception of candidate explanations it is much easier to comprehend the transition from an unusual experience of one's own thought to a delusional belief that some other entity inserted that thought. Plausibly, anyone who thinks that thought insertion is a nearby possibility and then subsequently has a highly irregular experience of her own thought would naturally attribute the thought to some other entity. We might wonder why the subject selects the specific entity she does rather than some other individual, but that will likely be determined by the contextual salience of certain individuals, as, for example, we can see in Saks's attribution of inserted thoughts to the houses on the street. Therefore I think we are in a good position to understand how, when one's experience has the structure of (A) and (B), irregular cognitive processes could generate a delusional belief in thought insertion. Furthermore, as I have argued, so long as these processes involve more than simply endorsing or accepting the first-person content of (A), we do not have to reject the Immunity Thesis.

[31] Naturally this raises the question of why a delusional subject would develop such an odd conception of possibility, a question that deserves much more attention than can be allotted for it in this essay. I discuss this issue more fully in Parrott, forthcoming.

5. Prediction Error Theory

But why would a person have a sense that her ordinary first-person awareness of her own thought was different or strange? Why would (B) be true? We ultimately want a much better understanding of the causes that generate the sense of abnormality in (B). Although answering this question completely would exceed the limits of this essay, in this section I would like to suggest one promising avenue of explanation.

It has been fairly well established that some type of self-monitoring is responsible for keeping track of bodily actions. Different neurocomputational theories have been developed to formally illustrate how this type of system might work. For example, in earlier work, Frith (1992) suggested that there is a single mechanism responsible for self-monitoring, but in more recent work he and colleagues posit multiple 'comparator' mechanisms (cf. Blakemore, et. al. 2002; Frith, 2005). The notion of comparator mechanisms has become a highly influential research paradigm in the cognitive sciences. The basic way such a mechanism is thought to work is by comparing a predictive model of the sensory consequences of a motor instruction with actual sensory feedback one receives. To give a simple illustration, suppose that I wave my hand in front of my face. The motor instruction to move my hand will generate a prediction of a certain pattern of visual feedback. When my visual system registers my hand actually waving in front of my face, this stimulus can be compared to the predicted consequences of the motor instruction (Blakemore, 2003). If I see what was predicted, there is a kind of 'match' between the actual visual stimulus and the prediction (which need not be exact). If not, there is a 'mismatch' between the two and my hand waving in front of my face will seem to be a highly unusual event, which means it will seem to be not self-generated, or like an 'unpredicatable passive movement' (Frith and Johnstone, 2003, pg. 138). This basic

framework has been used to explain, among other things, delusions of alien-control and auditory hallucinations (Blakemore, et. al. 2002; Frith, 2005; Frith, 2012; Ford, et. al. 2007; Fletcher and Frith, 2009).

Comparator mechanisms have also been thought to be the source of the 'sense of agency' that one has for bodily actions, in the sense that whenever there is a match between the predictive model and actual feedback, bodily movements are tagged with a 'sense' or 'feeling' of agency.[32] However, in his most recent work on this issue, Frith appeals to a predictive coding approach that emphasizes how self-monitoring is accomplished primarily by the production of prediction error signals when predictive models fail to match sensory stimuli (Adams, et. al., 2013; Corlett, et al*.,* 2009, and Fletcher and Frith, 2009). According to this framework, error signals are the driving force of computational processing. So, instead of thinking of comparator mechanisms as somehow tagging self-generated movements with a positive 'sense' or 'feeling' of agency, we need only think of them as tagging unpredicted movements with error signals. Applied to psychiatric phenomena, the idea would be that a person has a sense that her self-generated actions are not the result of her own agency because, when the consequences of her action are unpredicted, they generate a prediction error signal. As Fletcher and Frith describe it, 'a disruption in the prediction error accompanying self-generated actions could lead to those actions being felt as strange and externally generated.' (2009, pg. 55)

The model presented in this essay naturally suggests that this sort of prediction error theory could be extended to thought insertion. The central hypothesis would be that a person has an unusual experience of her thought (B) when it is not predicted. This could

[32] For additional criticism of the idea that comparators generate a positive feeling of agency, see Grunbaum, 2015.

happen, for instance, if there were some kind of impairment to whatever system generates predictive models of the consequences of her cognitive actions. As a result, there would be unpredicted feedback, which would generate an error signal, and the person would plausibly experience her episode of thinking as something highly abnormal or anomalous.

One problem with this way of developing the prediction error theory is that thinking, unlike bodily action, does not seem to involve any prior intentions or motor instructions. But without one of these, there does not seem to be anything to generate a predictive model of the expected consequences of thinking θ (c.f. Gallagher, 2004; Pacherie, et al., 2006; Vosgerau and Voss, 2014). One might try to respond to this concern by claiming that the predictions of cognitive actions are generated at a sub-personal level (cf. Campbell, 1999), but it is very hard to see what the purpose of such a mechanism would be. Unlike bodily action, it does not seem that a cognitive system has to keep track of the performance of one's mental actions in order to successfully think something. So even if the prediction error framework can help us explain delusions with respect to bodily actions, it is doubtful that a precisely analogous explanation can be given for thought insertion (cf. Frith, 2012).

There is, however, a different way in which the prediction error framework could explain the sense in (B). Rather than thinking subjects have some kind of deficit in a system that generates predictive models, we might instead hypothesize that a dysfunctional mechanism spontaneously generates aberrant prediction error signals in a manner that codes conscious thoughts as unpredicted or irregular. On this sort of view, the error signal would be autonomously generated by some kind of low-level impairment, rather than resulting from a problem with one's predictive model. Moreover, this sort of impairment would generate aberrant error signals in cases where one would not ordinarily expect to find them, namely for episodes of conscious thinking. Thus, it may be that a significant reason why an

experience of thought insertion seems so extremely salient to an individual is that conscious thoughts are not even the kind of thing that would normally be tagged with an error signal. Although this is a somewhat speculative hypothesis, which is in need of further empirical confirmation, we do have some corroborating evidence insofar as irregularities in error signaling are correlated with the pattern of dopamine firing we find in cases of schizophrenia (cf. Corlett, et. al. 2009; Fletcher and Frith, 2009; Frith 2012; Kapur, 2003).

This brief discussion of prediction error theory illustrates a plausible explanation of (B) might plausibly proceed. But the model presented in the earlier sections of this essay does not depend on it. We may ultimately need to appeal to some other kind of cognitive mechanism to explain (B), but I think the occurrence of a sense or feeling as if one's experience is highly unusual is something that cognitive science has the resources to explain.

6. Conclusion

This essay presents a model of both the experience thought insertion and the pattern of reasoning by which a subject undergoing such an experience comes to believe that her thoughts belong to some other entity. However, even if this model is right, additional questions must be addressed before we have a complete understanding of the phenomenon. Most importantly, nothing that I have said in this essay sheds any light on why a subject's belief that some entity is inserting thoughts persists in the face of counterevidence. Although it may seem obvious that the cognitive processes which normally sustain belief are impaired in cases of thought insertion, it is not clear precisely how they are different from those found in ordinary cognition. Nonetheless, I do not think we need to understand why a subject's belief in thought insertion persists in order to understand why it is adopted. Although it is

possible that future work will disconfirm the model presented in this essay, I think it currently offers the most plausible account of what it is like for someone to experience thought insertion.

Department of Philosophy

King's College London

References

Abramowitz, J. S., S. Taylor, & D. McKay 2009: Obsessive-compulsive disorder. *The Lancet*, 374: 491-499.

Adams, R. A., K. E. Stephan, H. R. Brown, C. D. Frith, K. Friston 2013: The computational anatomy of psychosis. *Frontiers in psychiatr*y 4.

Bayne, T. 2010: *The Unity of Consciousness*. Oxford: Oxford University Press.

Bayne, T. and Pacherie, E. 2004: Bottom-up or top-down?: Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry, and Psychology*, 11: 1-11.

Bayne, T. and Pacherie, E. 2005: In defense of the doxastic conception of delusions. *Mind and Language*, 20: 163-188

Blakemore, S. 2003: Deluding the motor system. *Consciousness and Cognition*, 12: 647-655.

Blakemore, S., Wolpert, D. and Frith, C. 2002: Abnormalities in the awareness of action. *Trends in Cognitive Science*, 6: 237-242.

Bortolotti, L. 2010: *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.

Bortolotti, L. and Broome, M. 2009: A role for ownership and authorship of thoughts in the analysis of thought insertion. *Phenomenology and the Cognitive Sciences*, 8: 205-224.

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

Campbell, J. 2001: Rationality, meaning and the analysis of delusion. *Philosophy, Psychiatry and Psychology*, 8: 89-100.

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

Carruthers, G. 2012: The case for the comparator model as an explanation of the sense of agency and its breakdowns. *Consciousness and Cognition*, 21: 30-45.

Clark, A. 2013: Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36: 181-204.

Coliva, A. 2006: Errors through misidentification: some varieties. *Journal of Philosophy*, 103: 407-425.

Coliva, A. 2002: Thought insertion and immunity to error through misidentification. *Philosophy, Psychology, and Psychiatry*, 9: 27-34.

Coltheart, M., Langdon, R. and McKay, R. 2011: Delusional belief. *Annual Review of Psychology*, 62: 271-298.

Coltheart, M. 2007: The 33rd Sir Frederick Bartlett lecture: cognitive neuropsychiatry and delusional belief. *The Quarterly Journal of Experimental Psychology*, 60: 1041-1062.

Coltheart, M. 2005: Conscious experience and delusional belief. *Philosophy, Psychiatry and Psychology*, 12: 153-157.

Corlett, P., Krystal, J., Taylor, J. and Fletcher, P. 2009: Why do delusions persist? *Frontiers in Human Neuroscience*, 3: 1-9.

Currie, G. 2000: Imagination, delusion and hallucinations. *Mind and Language*, 15: 168-183.

Davies, M., Coltheart, M., Langdon, R. and Breen, N. 2001: Monothematic delusions: towards a two-factor account. *Philosophy, Psychiatry and Psychology*, 8: 133-158.

Davies, M. and Egan, A. 2013: Delusion: cognitive approaches, Bayesian inference and compartmentalization. In K.W.M. Fulford, M. Davies, R.G.T. Gipps, G. Graham, J. Sadler,

G. Stanghellini and T. Thornton (eds.), *The Oxford Handbook of Philosophy of Psychiatry*. Oxford: Oxford University Press.

Doron, G. and M. Kyrios 2005: Obsessive compulsive disorder: a review of possible specific internal representations within a broader cognitive theory. *Clinical Psychology Review*, 25: 415-432.

Ellis, H. and Young, A. 1990: Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157: 239-48.

Evans, G. 1982: *The Varieties of Reference*. Oxford: Oxford University Press.

Fernandez, J. 2012: *Transparent Minds: A Study of Self-Knowledge*. Oxford: Oxford University Press.

Fine, C. M. Gardner, J. Craigie, & I. Gold 2007: Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry*, 12: 46-77.

Fletcher, P. and C. Frith. 2009: Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nature Reviews Neuroscience*, 10: 48-58.

Ford, J. Roach, B, Faustman, W., and Mathalon, D. 2007: Synch before you speak: auditory hallucinations in schizophrenia. *American Journal of Psychiatry*, 164: 458-466.

Frith, C. 2012: Explaining delusions of control: the comparator model 20 years on. *Consciousness and cognition*, 21: 52-54.

Frith, C. 2005: The self in action: lessons from delusions of control. *Consciousness and cognition*, 14: 752-770.

Frith, C. 1992: *The Cognitive Neuropsychology of Schizophrenia*. East Sussex: Psychology Press.

Frith, C. and E. Johnstone 2003: *Schizophrenia: A very short introduction*. Oxford: Oxford University Press.

Gallagher, S. 2007: Sense of agency and higher-order cognition: levels of explanation for schizophrenia. *Cognitive Semiotics*, 1: 33-48.

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

Hoerl, C. 2001: On thought insertion. *Philosophy, Psychiatry and Psychology*, 8: 189-200.

Hohwy, J. 2013: *The Predictive Mind*. Oxford: Oxford University Press.

Ismael, J. 2012: Immunity to error as an artifact of transition between representational media. In S. Prosser and F. Recanati (eds.), *Immunity to Error Through Misidentification: New Essays*. Cambridge: Cambridge University Press.

Jaspers, K. 1963: *General Psychopathology*. J. Hoenig and M. W. Hamilton (trans.), Manchester: Manchester University Press.

Julian, D., K. O' Connor and F. Aardema 2007: Intrusive thoughts, obsessions, and appraisals in obsessive-compulsive disorder: a critical review. *Clinical Psychology* Review, 27: 366-383.

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

Kuhn, W., Davis, M. H., and Lippman, S. 1988: Emotional adjustment to cardiac transplantation. *General Hospital Psychiatry*, 10: 108-113.

Maher, B. 1974: Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30: 98-113.

Maher, B: 1999: Anomalous experience in everyday life: its significance for psychopathology. *The Monist*, 82: 547-570.

O'Shaughnessy, B. 2000: *Consciousness and the World*. Oxford: Oxford University Press.

Pacherie, E., Green M. and Bayne, T. 2006: Phenomenology and delusions: who put the 'alien' in alien control? *Consciousness and Cognition*, 15: 566-577.

Parrott, M. forthcoming: Bayesian models, delusional beliefs, and epistemic possibilities. *British Journal for the Philosophy of Science*.

Peacocke, C. 2014: *The Mirror of the World*. Oxford: Oxford University Press.

Peacocke, C. 2008: *Truly Understood*. Oxford: Oxford University Press.

Peacocke, C. 1998: Conscious attitudes, attention, and self-Knowledge. In C. Wright, B. Smith, and C. MacDonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.

Pedrini, P. forthcoming: Rescuing the "loss-of-agency" account of thought insertion. *Philosophy, Psychiatry and Psychology*.

Pickard, H. 2010: Schizophrenia and self-knowledge. *European Journal of Philosophy*, 6: 55-74.

Recanati, F. 2012: Immunity to error through misidentification: what it is and where it comes from. In S. Prosser and F. Recanati (eds.), *Immunity to Error Through Misidentification: New Essays*. Cambridge: Cambridge University Press.

Roessler, J. 2013: Thought insertion, self-awareness and rationality. In K.W.M. Fulford, M. Davies, R.G.T. Gipps, G. Graham, J. Sadler, G. Stanghellini and T. Thornton (eds.), *The Oxford Handbook of Philosophy of Psychiatry*. Oxford: Oxford University Press.

Rowa, K., C. Purdon, L. Summerfeldt, & M. Antony 2005: Why are some obsessions more upsetting than others? *Behaviour Research and Therapy*, 43: 1453-1465.

Rowa, K. and C. Purdon 2003: Why are certain intrusive thoughts more upsetting than others? *Behavioural and Cognitive Psychotherapy*, 31: 1-11.

Saks, Elyn. 2007: *The Center Cannot Hold*. New York: Hyperion.

Shoemaker, S. 1970: Persons and their pasts. In John Perry (ed.), *Personal Identity*. Berkeley: University of California Press.

Sims, A. 2003: *Symptoms in the Mind*, London: Elsevier.

Sollberger, M. 2014: Making sense of an endorsement model of thought insertion. *Mind and Language*, 29: 590-612.

Soteriou, M. 2013: *The Mind's Construction*. Oxford: Oxford University Press.

Stephens, G. L. and Graham, G. 2000: *When Self-Consciousness Breaks: Alien Voices and Inserted Thoughts*. Cambridge: MIT Press.

Sousa, P. and L. Swiney 2013: Thought insertion: abnormal sense of thought agency or thought endorsement? *Phenomenology and the Cognitive Sciences*, 12: 637-654.

Stone, T. and Young. A. W. 1997: Delusions and brain injury: the philosophy and psychology of belief. *Mind and Language*, 12: 327-364.

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

Vosgerau, G. and Voss, M. 2014: Authorship and control over thoughts. *Mind and Language*, 29: 534-565.