

SCHIZOPHRENIA AND THE EPISTEMOLOGY OF SELF-KNOWLEDGE

HANNA PICKARD

Oxford Centre for Neuroethics and All Souls College, University of Oxford

ABSTRACT

Extant philosophical accounts of schizophrenic alien thought neglect three clinically significant features of the phenomenon. First, not only thoughts, but also impulses and feelings, are experienced as alien. Second, only a select array of thoughts, impulses, and feelings are experienced as alien. Third, empathy with experiences of alienation is possible. I provide an account of disownership that does justice to these features by drawing on recent work on delusions and self-knowledge. The key idea is that disownership occurs when there is a failure of rational control over one's mind. This produces a clash between the deliverances of introspection and practical enquiry as ways of knowing one's mind. This explanation places disownership on a continuum with more common aspects of our psychological life, such as addiction, akrasia, obsessional thinking, and immoral, selfish or shameful thoughts. I conclude by addressing objections, and exploring the relevance of my account to questions in the philosophy of psychiatry concerning the validity of our current taxonomy of symptoms, and the nature of psychiatric classification

Keywords: agency, alien thought, classification, delusions, disownership, empathy, introspection, practical enquiry, rational control, schizophrenia, self-knowledge, thought insertion

Over the past decade, there has been a flurry of philosophical interest in delusions in general; and one striking schizophrenic delusion, known as alien thought or, alternatively, as thought insertion, in particular.¹ Schizophrenics typically suffer from a range of symptoms, including severe social impairment or withdrawal; passivity or deadening of action, emotional expression, and speech; incoherence and disorganization in thought and speech; and also a range of delusions, especially paranoid or persecutory beliefs, self-referential and magical thinking, hearing voices, feeling controlled by outside forces, and, finally, alien thought (APA 1994). Alien thought is a first-rank, diagnostically central symptom of schizophrenia. It is attributed when schizophrenics report that they have conscious mental events that are not their own. To get an initial flavour of the symptom, consider the following well-known example of a patient's

¹ See, for example, *The Monist* volume 40 (1999) dedicated to Cognitive Theories of Mental Illness; volume *Philosophy, Psychiatry and Psychology* volume 8 (2001) dedicated to Understanding and Explaining Schizophrenia; *Mind and Language* volume 15 (2000) dedicated to Pathologies of Belief; as well as Bayne and Pacherie (2005); Bortolotti (2009); Bortolotti and Broome (2009); Fernandez (2010).

report: “Thoughts are put into my mind 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, 66)

But despite the flurry of interest, the philosophical literature on alien thought is characterized by a series of striking omissions. The first is the paucity of the diet of clinical examples: the same patient reports are used again and again, and they do not fully capture the phenomenon. The second is the failure to apply the most promising general account of delusions, namely, the two-factor model, to alien thought. The third is the isolation of philosophical discussions of alien thought from wider issues in philosophy of psychiatry, such as the possibility of empathy with patients, and questions of the validity of psychiatric classification. The upshot of these omissions is that it is often left unclear what, particularly from a philosophical point of view, needs explaining about alien thought, and how it connects to wider questions about mental health.

In this paper, I aim to offer an account of alien thought that addresses these omissions. The paper has five parts. I begin by enriching the diet of clinical examples of alien thought, and applying the two-factor model of delusions, in order to clarify what needs explaining. Secondly, I turn to the research that first brought alien thought to the attention of philosophers: the psychologist Chris Frith’s information-processing model of alien thought (1992) and the use made of it by John Campbell in his seminal paper ‘Schizophrenia, the space of reasons, and thinking as a motor process’ (1999). Campbell accepts Frith’s account, but links it to a novel and additional idea about ownership. I argue against the plausibility of Frith’s account, but for the importance of Campbell’s additional idea. Thirdly, I develop a model of self-knowledge that draws on Richard Moran’s account in his *Authority and Estrangement* (2001) and connect it to Campbell’s additional idea. The suggestion that Moran’s work has something to offer our understanding of alien thought is not new: both Jordi Fernandez (2010) and Lisa Bortolotti and Matthew Broome (2009) have employed it. But its connection to questions in the ontology of mind, and, correspondingly, its explanatory power with respect to the full range of clinical examples of the symptom, has not been sufficiently appreciated. So fourthly, I use that model of self-knowledge to develop an account of alien thought that addresses what needs explaining, and I consider some objections. Finally, I connect the discussion of alien thought to wider questions in psychiatry, about the possibility of empathy with patients, and the nature of psychiatric classification.

1. Alien thought: what needs explaining?

Philosophical examples of alien thought tend to come in one variety. The kind of mental event that is disowned by patients is a conscious, occurrent thought, like the injunction to ‘Kill God’ quoted above. But schizophrenics do not only disown thoughts. They also disown impulses and feelings. Consider the following patient

reports:

The sudden impulse came over me that I must do it [empty his urine bottle over the ward dinner trolley]. It was not my feeling, it came from the x-ray department . . . It was nothing to do with me, they wanted it done. (Mellor 1970, 17).

I cry, tears roll down my cheeks and I look unhappy, but I have a cold anger because they're using me in this way, and it's not me who's unhappy, but they're projecting unhappiness onto my brain. They project upon me laughter, for no reason, and you have no idea how terrible it is to laugh and look happy and know it's not you, but their emotions (Mellor 1970, 17).

In the first of these patient reports, the conscious, occurrent mental event that is experienced as alien is an impulse – indeed, an impulse to perform a wrongful and arguably aggressive act. In the second, it is a feeling of happiness or unhappiness, and the various forms of expressions those feelings take. Given these patient reports, alien thought is misnamed. There can also be alien impulses and feelings. At first glance, then, an account of alien thought should apply not only to thoughts, but also to impulses and feelings.

Consider next the following report by Elyn Saks:

As I walked along, I began to notice that the colors and shapes of everything around me were becoming very intense. And at some point, I began to realize that the houses I was passing were sending messages to me: *Look closely. You are special. You are especially bad. Look closely and ye shall find. There are many things you must see. See. See.*

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, 27).

Standard examples of patient reports attribute the alien mental event to another person or group of people. Saks here attributes her alien ideas to nearby houses. This highlights forcefully the need to distinguish two components of alien symptoms. On the one hand, there is a mental event that is experienced by the schizophrenic as not their own. On the other, there is the attribution of this mental event to another person or thing. In the above patient reports, these objects of attribution include the chap Chris, the x-ray department, a vague and ominous 'they', and the nearby houses.

These two components can be helpfully distinguished by considering the most promising general model of delusional belief: the two-factor model. The two-factor model has been developed in a series of papers by Martin Davies, Max Coltheart, and various of their colleagues (Davies and Coltheart 2000; Davies, Coltheart et al. 2001; Aimola Davies and Davies 2009). But the basic idea is straightforward. Delusions depend on at least two factors. On the one hand, there is an aberrant or unusual experience that is the ground for the delusional belief. On the other hand, there are the various belief formation and retention processes that cause the wayward experience to be taken as grounds for belief in the first place, and the belief to be retained, even in the face of strong counter-evidence. Applying this model to alien thought is helpful: it allows us to isolate what the basic content of the aberrant or unusual experience is likely to be, as against the more cognitive processes likely to be involved in the delusion.

Consider again the range of objects of attribution in the above patient reports. Their variety, idiosyncrasy, and detail, makes it highly implausible to think that the attributional component of alien thoughts, impulses, and feelings could be part of the initial, wayward experience. How could the first factor of the delusional process be an experience that presents a mental event as belonging to Chris, let alone to the distant x-ray department, 'they', or some nearby houses? For, in some of these cases, the object itself is not even experienced at the time of the delusion; in others, although experienced, it is not evident, to say the least, that or how it could be experienced as possessing the mental event in question. Indeed, that mental event seems to be experienced as possessed by the schizophrenic. Rather, the first factor seems more plausibly simply to be an experience of having a mental event, whether a thought, impulse, or feeling, that does not seem to be one's own. The attribution of this mental event to another person or thing can then be explained by cognitive processes constitutive of the second factor. For instance, there is evidence that both schizophrenic and paranoid patients show a generalized attributional reasoning bias towards assigning causal responsibility for events to others, rather than to assigning them to themselves (Baker and Morrison 1998; Moritz 2007). It is thus natural to hypothesize that, given that a schizophrenic has a thought, impulse or feeling they experience as not their own, they would be likely not only to believe that it is not their own, but, further, to believe that it belongs to another. Which particular person or other is then selected for the attribution will in all likelihood be context-dependent and patient-specific, determined by factors such as the patient's history, together with current pre-occupations and present actual situation. In effect, in face of the wayward experience of alienation, we can see the schizophrenic's delusion that it belongs to another as offering a paranoid inference to the best explanation: if it's not theirs, it must be someone or something else's.

Hence by employing the two-factor model, we can discern two layers to alien thought. There is, in the first instance, what we might think of as a minimal experiential content: the experience of having a mental event, whether a thought,

impulse, or feeling, that does not seem to be one's own.² That is the first factor. Then there is the attribution of the mental event to another. That is the second factor: it depends on the various cognitive processes involved in the formation and retention of this kind of delusional belief.

This clarification is important for two reasons. First, because, together with a richer diet of patient reports, it delineates what needs to be explained. What needs to be explained is how it is possible to have a mental event that does not seem to one to be one's own. We need a philosophical account of this minimal experiential content.

Second, the clarification is important because it both suggests and helps to explain the possibility of empathy with schizophrenic patients. Part of what is so striking about alien thought is its combination of extreme mental disturbance with comprehensibility. We do not treat alien thought as incoherent: for example, it is not part of the cluster of symptoms involving disorganized thought or speech. We take the schizophrenic as reporting their experience accurately, and saying something meaningful. This suggests that, despite the deep irrationality expressed in the above patient reports, we do seem to have some idea what the pathology is like: to some extent, at least, we can empathize with, and not just pity, the schizophrenic suffering from alien thought. Such empathy can potentially be explained by recognizing the minimal experiential content of alien thought. For this minimal experiential content seems not uncommon. There exist many aspects of our psychological lives that, although less disturbed than alien thought, seem potentially to lie on the same continuum. Addiction, akrasia, obsessional thinking, immoral or selfish or shameful thoughts – in describing our relationship to these aspects of our minds, we reach for the idea of failures of ownership and identification. We often wish to dissociate ourselves from them: they are not truly our own. In contrast, the idea of having a mental event like a thought, impulse, or feeling, that one takes actually to belong to some other person or thing who has put it into one's mind, is certainly much less ordinary. Indeed, it tends towards the incoherent if too closely scrutinized. I do not say we can make no sense of it whatsoever. After all, paranoid and magical thinking is not confined to pathology. But recognition of the layered structure of alien thought helps to explain our capacity to feel we have some grip on the phenomenon and to empathize with patients: it is our foothold into this aspect of the schizophrenic's world.

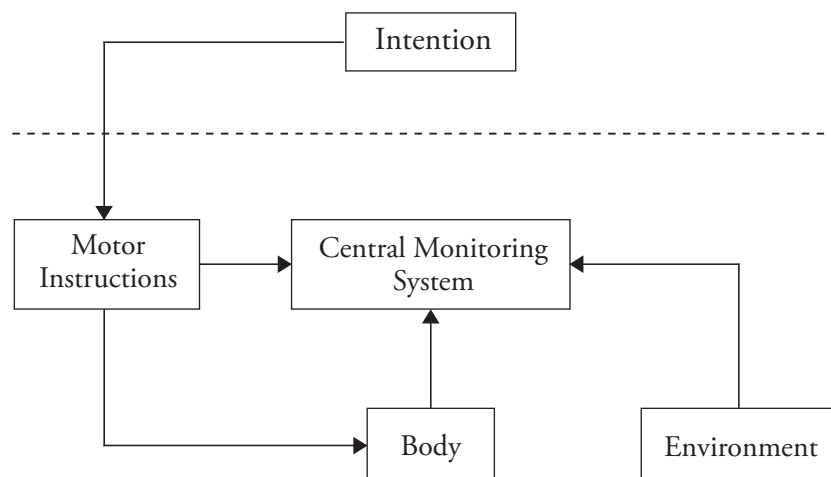
2. Thinking as a motor process

How is it possible to have a mental event that one experiences as not one's own? The origin of the recent philosophical interest in alien thought is Chris Frith's account.

² Note that this is to be distinguished from not experiencing a mental event as one's own. Arguably, that is the correct description not of alien thought, but of the experience of subjects, like animals and infants, that lack self-consciousness.

This account employs an information-processing model of ordinary thought that treats thinking as a motor process: thought as a kind of mental action (1992). So consider first the core idea of the relevant information-processing model of action. Having decided to pick up a cup that one sees, how does one do it? Normally, there is no personal-level explanation available. From one's own point of view, one just does it: one just picks it up. However, there is a sub-personal-level explanation. Sub-personal processes compute the type of trajectory, finger grasp, and so on, that the action requires, and a series of instructions for movement are sent to the limb in question. But there is strong evidence that copies of these instructions are also sent to a central monitoring system. This system also receives internal feedback from the body, and perceptual information about the environment. It can then compare motor instructions, internal feedback, and perceptual information.³ Very crudely, we can diagram the model thus:

Cognitive Model of Action



This model has at least two advantages. First, by comparing the copy of instructions for movement with perceptual information about the environment, the system is able to compute that the movement needs to be corrected – say, because the target has moved – or compute what the next, compensatory movement should be, before the initial movement has been completed, and internal feedback from the body to this effect received. In essence, the model explains the speed and efficiency of action.

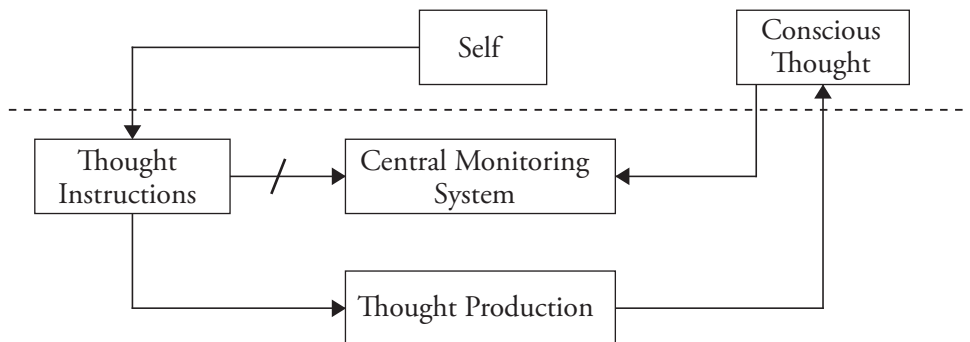
Second, by comparing the record of instructions for movement with the internal feedback from the body, the system has a way of checking whether or not they match. This not only allows it to tell whether an instruction has in fact been executed. It also allows it to distinguish between bodily movements which are intended, and bodily movements which are not. The prospects of explaining alien thought by treating thinking as a comparable motor process depend crucially on this last point.

³ See Bosgerau and Newen (2007) for a discussion of the computational demands of this comparison. The model was initially developed by von Holst and Mittelstaedt (1950) and Sperry (1950).

Neuropsychologists conjecture that our sense of agency – the experience of a bodily movement as willed or under our control – must somehow result from a match in the central monitoring system.

Frith proposes that we can explain the experience of alien thought by positing a comparable central monitoring system for thought. Clearly, there must be some sub-personal explanation of how a conscious thought is produced by the brain. So suppose that, normally, sub-personal processes not only issue instructions for the production of a particular thought, but also send copies of these instructions to a central monitoring system. Suppose, too, that, once the thought has been produced, the central monitoring system also receives information to that effect. We can now explain alien thought by hypothesizing that, in these cases, the central monitoring system does not receive a copy of the instruction to produce a particular thought. The central monitoring system would then receive information that a particular thought has been produced, of which it has no record of an instruction for its production. Frith's conjecture is that such a mismatch would result in the schizophrenic experience of disownership. The thought would be experienced as alien because it would not be experienced as produced by the self: there would be no sense of agency. As Frith puts the point: "It is as if each thought has a label on it saying 'mine'." (1992, 80) Otherwise put, the suggestion is that there is a phenomenal aspect to experience – a sense of ownership – that results from a match in the central monitoring system. This sense of ownership normally characterizes conscious thought, but is missing in the schizophrenic case. That is why the thoughts are experienced as alien. We can diagram the model thus:

Frith's Cognitive Model of Thought



There are three common objections to this model. The first is that it leads to an infinite regress. The thought is that there must be an intention to produce a thought, whether that intention is conscious or not. But an intention is itself a thought, hence there must be an intention to produce the intention to think a thought, and so on ad infinitum (Gallagher 2004; Vosgerau and Newen 2007).

This objection depends on one of two faulty assumptions. If we suppose the intention is conscious, then it is not correct to treat it as a thought. Intentions are standing states, part of one's background psychology, and so similar to beliefs, desires, and emotions. They can thus be manifest in one's stream of consciousness, for instance, as a thought, but they are not themselves identical with such manifestations. Therefore they are not the kind of thing which Frith's model aims to explain the production of. Alternatively, if we suppose that the intention is unconscious, then it is common for the objection to identify it with the motor instructions to produce a thought. Once identified, we now seem to need to explain how a 'motor intention' is produced – a kind of sub-personal thought. But such an identification is also simply a mistake. It expresses a slide from one idea to another, and unjustifiably imports the idea of a personal-level psychological state, with all that entails, to sub-personal processes. There is no good reason to conceive of motor instructions – for thought or action – as intentions.

The second objection to this model is that it cannot explain why a lack of match at the central monitoring system should result in the experience of a thought as alien, as opposed, more simply, as not experienced as intended (Marcel 2003). It may be helpful to compare thought with bodily movements to see the point of this objection. Expressive gestures, for instance, are not actions. Indeed they are often initiated automatically by lower brain systems. So there will be no match at a central monitoring system for action. But they are not thereby experienced as not one's own. They are simply not experienced as intended: they are not actions. Why should thought be any different? For mental activity is not always intentional. Our minds wander idly. We daydream and fantasize. Having a thought enter your mind unbidden, unexpected, and even unwelcome, is a perfectly ordinary part of mental life. Hence the objection is that the model does not explain why a lack of match in the central monitoring system would cause a thought to be experienced as not one's own, as opposed, more simply, to be experienced as one's own, but not as intentionally produced.

This objection depends on presuming too close a similarity between the production of thoughts and the production of bodily movements. For, as the objection itself points out, not all actions are initiated by the motor cortex; expressive gestures are usually initiated automatically by lower brain systems. In contrast, Frith's model is supposed to explain how *all* thoughts are produced – whether or not these are part of an intentional mental activity, like solving a problem, or an idle daydream or fantasy. Perhaps this breadth of scope is mistaken; perhaps different thoughts are produced by different systems. But, once it is acknowledged that the model is supposed to encompass all thoughts – the more as well as the less intentional – then it is intuitive that a lack of match at the central monitoring system would cause one to experience the thought as not one's own. For one's mind would contain a thought, for which one has no record of an instruction for its production. So where does it come from?

The third and final common objection to this model is that it is difficult to see what its point or purpose could be: how could it have evolved? For it is hard to see an evolutionary need for a system capable of distinguishing, among thoughts one has, those which are one's own, from those which are not. However, it is possible that this capacity is an epiphenomenon of a more functional system. Campbell suggests that the point or purpose

of the system may be to keep track of one's thoughts, and to ensure that they cohere and relate to one's overall cognitive aims (1999).

Suppose that the central monitoring system has access not only to one's conscious thoughts, but also to one's overall cognitive aims, such as the exploration of a particular line of reasoning, or the construction of an argument designed to convince someone to do one's bidding. Just as a central monitoring system for action will contribute to speed and efficiency, so too then will a central monitoring system for thought. The system will not have to wait for the thought to have occurred and be recorded before determining whether or not it furthers one's aim: it can compute this using the copy of the instruction to produce the thought. And we do sometimes seem to experience something like this. Sometimes one realizes as one is starting to think a thought – perhaps even just before – that it is in some sense, relative to one's aim, mistaken.

However, attributing this function to the central monitoring system risks making thinking itself an epiphenomenon. For now it appears that the conscious thought itself is not necessary: the whole thought process can unfold and be kept on track sub-personally, and the very possibility of a self who is responsible for the production of his or her own thoughts is lost (Vosgerau and Newen 2007).

It is possible to question the force of this objection. On the one hand, there does seem to be something inherently passive and outside of oneself, as it were, about the process of thinking. Many accounts of creative genius and insight report the idea just coming in a flash (Wegner 2002). So, although the comparison with action may sometimes be apt, sometimes thoughts do not seem to be produced or controlled by the self in the way actions typically are. On the other hand, even if the process of thinking is passive in this way, there could still be a purpose to the production of conscious thought. It need not be merely epiphenomenal, because, even if the point of the central monitoring system is to keep track of one's thoughts, it might be the consciousness of the thought that allows one intentionally to communicate it, or to decide what to do with it next. But be that as it may, Campbell takes the risk seriously. Indeed, he develops Frith's account to deal with just this concern.

Campbell proposes that the self is still rightly conceived of as the producer of his or her own thoughts because it is his or her own standing, background beliefs, desires, and emotions which determine, together with external stimuli, which motor instructions for thoughts are issued. Treating thinking as a motor process allows us to begin to model how thoughts are produced, and to explain how or why the schizophrenic judges that a particular thought is alien: it lacks the sense of ownership that normally characterizes thought. But it need not strip the self of agency with respect to thoughts because the sub-personal system underlying the production of conscious thoughts is driven by the self's personal-level mental states. Hence Campbell develops Frith's account by linking the sense of ownership produced via sub-personal monitoring of the production of thoughts, with the idea of ownership as a form of dependency on standing, background beliefs, desires, and emotions: "This dependency of which thoughts you have on your underlying psychology has to do with our sense of ownership of thoughts: that the particular thoughts you have belong to you, rather than being shared by many people" (1999, 617).

This idea is intuitively credible. But note two things. First, part of the appeal of Frith's model is that it links ownership with agency: one has an experience of owning a thought just in case one has a sense of being its producer. At first glance, it is not obvious that Campbell's additional idea preserves this link. Second, although the idea helps salvage Frith's proposal that thinking is a motor process, it is wholly independent of it. In principle, it is compatible with whatever sub-personal model of the production of thoughts proves to be empirically correct.

This is important. I have suggested that the common objections levied against Frith's account can be met. But if one steps back, there are three broader considerations that should make us question the plausibility of the model, even if, unlike the aim of the more common objections, they do not refute it outright.

The first consideration is that the model applies only to thoughts, and not to impulses and feelings. It does not aim to explain how these are produced, let alone experienced as alien. But as I pointed out in the first part of this paper, schizophrenics disown not only thoughts, but also impulses and feelings. As it stands, it is not obvious how we could extend Frith's model of alien thought to alien impulses and feelings. We would thus be required to abandon the prospects of a unified account of the schizophrenic disownership of mental events.

Secondly, the model offers no explanation of why only a particular selection of thoughts are experienced as alien. For schizophrenics do not experience all their thoughts as alien. Nor, for that matter, does the selection seem random. For instance, the schizophrenic thinks that the thought 'Kill God' is alien. He or she does not think that other thoughts they may also have, such as 'The thought 'Kill God' is not my thought' or 'Grass is green' are alien. But why? Why would there be a breakdown in the mechanism for sending copies of instructions to central monitoring for only a very select array of thoughts, and not others? It is natural to try to answer this question not by consideration of sub-personal processes, but by consideration of the person. What is it about thoughts with these particular contents that would prompt a sense of alienation in this schizophrenic?

Thirdly and relatedly, the model makes the comprehensibility of the pathology and the possibility of empathy with patients utterly mysterious. Either there is a match at the central monitoring system, or there is not. For those of us who do not suffer from alien thought, the hypothesis must be that there is always a match: we never fail to experience a sense of ownership accompanying our conscious thoughts. We thus have no experience of what it is like for there not to be a match: for there to be no accompanying sense of ownership. But that is just to say we have no foothold whatsoever to help us understand what alien thought might be like – no possibility of a subjective grasp of the pathology. And, although we may struggle to fully understand the schizophrenic's world, it does not seem to be wholly alien. Frith's model does not allow for this connection between the schizophrenic and us. Nor, correspondingly, does it lend itself to the possibility of a continuum between mental health and disturbance.

3. An epistemological framework for alien thought

Bearing these considerations in mind, I want to propose an alternative account of alien thought. The account can be seen as an elucidation of Campbell's additional idea about ownership. Recall Campbell's suggestion: "What makes my occurrent thoughts mine is not just that they show up in my stream of consciousness. What makes them mine is, in addition, the fact that they are products of my long-standing beliefs and desires, and that the occurrent thinking can affect the underlying states" (1999, 621). Campbell's idea is that ownership involves a two-way causal dependency. Standing, background states causally affect occurrent mental events, and occurrent mental events causally affect standing, background states. So suppose that we put Frith's model to one side, and focus solely on this idea. The questions that arise are: Why would this two-way dependency be connected to experiencing a mental event as one's own? Is there a way of preserving the intuitive link between ownership and agency that Frith's account highlights? And, in absence of Frith's account, how can we explain the schizophrenic minimal experiential content: the experience of a mental event, whether a thought, impulse, or feeling, as not one's own?

To answer these questions, I want to draw on a distinction Richard Moran makes in his *Authority and Estrangement* (2001). This may initially seem like a diversion, but it is not. For Moran's overall aim in this book is to explore the way one is not just a passive subject of one's mental life, but rather its agent or author. And his key idea is that this contrast is evident in two different stances one can take to one's mind: the theoretical stance, and the practical stance.

The theoretical stance involves treating one's mind as a pre-existing realm of mental states and events, whose nature is there to be discovered. So one can ask oneself 'What *do* I believe, or desire, or feel?' on the assumption that *there is an antecedent fact of the matter* about one's beliefs, desires, or emotions. Typically we answer this sort of theoretical question through introspection: we look within. However, it is perfectly possible, if more rare, to answer it by considering instead one's physical behaviour: to take up a more third-person point of view on oneself.

The practical stance, in contrast, involves treating one's mind as something *to be made up* through the process of deliberation or reflection. Evidently this is only possible for some types of mental state: those which are responsive to reason. Usually one cannot, for instance, simply and directly make up one's mind about whether or not to feel pain, or about whether or not to perceive what is salient in one's perceptual field (although one can, of course, exercise some control over these states, e.g. one can choose to take painkillers, or to shut one's eyes). But one can, for instance, ask oneself 'What, given the circumstances, should I believe, or desire, or feel?' in the spirit of determining directly what *to* believe, or desire, or feel. And this question cannot be answered through introspection any more than it can be answered by considering one's physical behaviour. It is 'transparent' to the world, as it is sometimes put. To answer it one must consider what beliefs, desires, or emotions would be appropriate or warranted: one must consider what the world in which one finds oneself is actually like.

Moran is interested in the practical stance because he believes that it is a hitherto unrecognized source of self-knowledge. In essence, he claims that one can know what one believes, desires, or feels by determining what to believe, desire, or feel. This is a striking claim, but there are immediately two problems with it which Moran does not address. The first problem is that it is unclear how determining what to believe, desire, or feel by considering what *the world* is actually like could ever issue in knowledge of *the self*. On the face of it, the subject matter is simply wrong. I shall return to this problem shortly. The second is that, putting this first problem aside, the mere possibility of the practical stance yielding self-knowledge presupposes that the process of reflection *does in fact* make up one's mind: it has a causal effect. Otherwise it could never be a source of self-knowledge. And whether or not this presupposition is warranted depends on two things: the type of mental state in question, and the aim of the enquiry.

One aim of practical enquiry is the acquisition of new mental states. Suppose that, upon reflection, one judges that *p*. If one has no antecedent views on the matter, and one's memory is functioning normally, then it is nearly impossible for one to fail thereby to acquire the belief that *p*. For this reason, we might be inclined to allow that, at least *knowing* that one is judging that *p*, if not judging that *p* itself, can be a reliable method of knowing what one believes (Soteriou 2005). This is one way to explain how a judgment about the world could yield knowledge of the self: when one makes a judgment about the world, one knows that one is making that judgment, and so one has knowledge of the self. Moran himself may not welcome this suggestion, since it undermines the idea that the practical stance is a source of self-knowledge independently of the theoretical stance. But it does preserve a role for practical enquiry in the epistemology of self-knowledge. However, it also highlights a contrast between beliefs on the one hand, and desires and emotions on the other. For, it is relatively common to judge that an object is desirable, say, or that a situation warrants a particular emotional response, and yet not thereby to acquire the appropriate desire or emotion. Knowing that one is making such a judgment is therefore not such a reliable method of knowing what one wants or feels.

But practical enquiry can also aim at the revision or maintenance of a pre-existing mental state. And when it does, the difference between belief, desire, and the emotions, is much less pronounced. Suppose that one makes a judgment which undermines a pre-existing belief, desire, or emotion. If one is not conscious of having the mental state in question, then it is possible that it will not be abandoned or revised. For instance, a person may consistently judge that racism is wrong, say, while their behaviour makes it evident that they continue to hold various beliefs, desires, and emotions, which conflict with this judgment. So long as these mental states are all underground, as it were, they are all equally impervious to revision. But if the mental state which the judgment undermines is conscious – it is in no way repressed – then irrespective of the type of mental state it is, it is much more rare for it *not* to be abandoned or revised. On the whole, one cannot consciously believe that *p* after judging that not-*p*. Nor can one continue to be consciously angry at *X* for committing a crime, say, after judging *X* innocent. I do not claim that we can make no sense of conscious mental states being resistant to such revision. There are certainly such things as irrational

and immovable phobias, or forms of perversity, or more simply, quirks of human psychology, for instance, it can sometimes simply take time for emotions to subside. But, in general, such resistance to revision is not the norm, and it is usually, if not always, expressive of a sort of mild irrationality if not pathology. For this reason, it seems that knowing that one is making a judgment that undermines a conscious mental state *is* a reliable method of knowing that one no longer has that state. It is not infallible. But it is reliable.

So we have a kind of control over our mental states in so far as their acquisition, and even more, their maintenance and revision, can in fact be determined by a process of practical deliberation or reflection. When joined with the theoretical stance and the deliverances of introspection, this process can yield self-knowledge. With this framework in mind, let us return now to alien thought.

4. An alternative account of alien thought

Schizophrenics do not disown mental states. Nor, we might add, do they disown perceptions and sensations. They disown only conscious thoughts, impulses, and feelings. Why? These form a unified set of mental phenomena. They are all mental events that can be conscious manifestations or expressions of mental states. We are familiar with the idea that physical behaviour can manifest mental states. But mental behaviour can manifest them too. For instance, one's belief that *p* can be consciously manifest in a thought that *p* – a thought which can be triggered, most basically, by being asked the question whether *p*, but which can also be triggered by more random associations. Alternatively, one's standing jealousy of *X*, say, can be manifest in jealous pangs, or in ugly thoughts and impulses – feelings, thoughts, and impulses which can be triggered by encountering *X*, or again by more random associations. In this way, conscious thoughts, impulses, and feelings bring the mental states that they manifest themselves to consciousness: one can be conscious of one's beliefs, desires, and emotions by having conscious thoughts, impulses, and feelings that manifest them.

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. It is neither possible nor acceptable to 'Kill God'. It is not desirable to pour one's urine over one's ward dinner trolley. There is no evident reason for laughing hysterically or feeling overjoyed. Normally such judgments would result in the loss of the mental state in question. Normally one would then cease to think about killing God, to want to pour the urine, or to feel overjoyed. Correspondingly, normally one would expect all manifestations of these states to abate. In the case of schizophrenics, my suggestion is, the manifestations do not abate. Looking within, they are consistently lumbered with them. There is thus a radical, persisting, disparity between the mental states that they believe that they have, in virtue of having made up their mind about the matter, and

the mental states that seem to be manifest in their consciousness.⁴ That is why they disown these manifestations – why they do not believe that these relentlessly occurring thoughts, impulses, or feelings are caused by their own mental states, and so expressive of their own background psychology.⁵

We can now explain why Campbell is right that the two-way dependency between mental states and mental events is connected to ownership. It is connected to ownership because, initial appearances notwithstanding, it is connected to agency. Through practical enquiry, one exercises control over one's psychological life. Reflection and deliberation can culminate in judgments that endorse or reject as unwarranted one's mental states. This can and normally does have a causal effect on the continued existence of those states. So although one cannot decide by fiat what to believe, desire, or feel, one can decide what it is correct or appropriate to believe, desire, or feel, and thereby determine, at least to a reasonable extent, which states of mind one has. In this way, one has some control over and responsibility for the shape of one's psychological life. This is why one can feel identified with, or alienated from, one's mental states and the events which manifest them. Identification occurs when a mental state is causally responsive to one's rational will. Alienation occurs when a mental state is not. It occurs when, instead, it lies outside of one's rational control: although one does not endorse it, it continues to exist, impervious to this fact, and despite oneself. Of course, the extremes of identification and alienation lie on a continuum: no doubt, the phenomenon comes in degrees. We can thus preserve the link between ownership and agency that makes Frith's model compelling, while rejecting the model itself.⁶

We can also now explain how, in absence of Frith's model, schizophrenics experience a thought, impulse, or feeling as not their own. The explanation again depends on the role of practical enquiry, not now in determining the nature and existence of one's mental states, but rather as basis for self-knowledge. My suggestion is that schizophrenics disown mental events that seem to be manifestations of mental states they do not believe they have. This blatantly forces the question: how then do they form their beliefs about which mental states they have? Evidently, schizophrenics are not using only introspective consciousness of thoughts, impulses, and feelings as a way of self-ascribing mental states. Otherwise there could be no disparity between their beliefs about which mental states they have, and which mental states seem to be manifest in their minds. Equally, it seems implausible that schizophrenics are using behavioural evidence as a means of self-ascription: they are not likely to be taking up such an objective, third-person point of view on themselves. But, once we

⁴ Cf. Bortolotti and Broome (2009) and Fernandez (2010). These authors also employ the practical stance to explain alien thought, but they do not link this stance, nor, correspondingly, the explanation, to the ontology of mind and the possibility of rational control over one's mental states.

⁵ Cf. Stephens and Graham (2000) who suggest that disownership involves a lack of 'interpretive mesh' between a mental event and one's psychological life.

⁶ Harry Frankfurt famously offers a similar if distinct account of alienation. He suggests that an addict who feels compelled to take a drug "may meaningfully make the analytically puzzling statement that the force moving him is other than his own" (1971, 13). Frankfurt's explanation of this is that the addict does not want to take the drug: his second-order desire is out of step with his first-order desire. The differences between Frankfurt's account and the proposal I have suggested are twofold. First, Frankfurt does not link second-order desires to practical enquiry into what one should desire. Second, and correspondingly, second-order desires are not, according to Frankfurt, causally potent. They simply offer their approval to some, but not all, first-order desires – much as a passive bystander could.

recognize that practical enquiry can be a source of self-knowledge, we can answer this question. For schizophrenics, there is in effect a clash between the deliverances of two first-personal methods of self-ascribing mental states: introspection and practical enquiry. The idea of a clash between ways of knowing one's own mind is familiar. Such a clash can exist between the deliverances of introspection, and the deliverances of third-personal evidence, when, for instance, a mental state is repressed or denied, and yet manifest in physical behaviour. In effect, I am suggesting that there can also be a clash which is more internal to the mind: internal to the first-person perspective. Through practical enquiry, schizophrenics come to believe that they do not have mental states which introspection on its own would lead them to self-ascribe. That is why it is possible for them to experience a mental event as not their own: it is a manifestation of a mental state that they do not, on other grounds, believe they have. This is why it feels so alien.⁷

There are two natural objections one might raise to this account of alien thought. The first is that the clash between first-personal ways of knowing one's own mind is not a sufficient explanation of the symptom. It is possible to think that this explanation likens the schizophrenic experience too closely to our own: in placing the experience on the continuum of alienation or failures of identification that characterize more mundane and common irrationality, we strip the phenomenon of its pathology. But, in response, recall that this explanation of minimal experiential content is simply the first factor in an account of delusional belief. The second factor, whereby various cognitive processes attribute the mental event experienced as alien to another person or object, are also part of the explanation. Moreover, it is arguable that, quite generally, to understand why any symptom is genuinely pathological requires placing it in a wider psychological context of mental disturbance (Wu 'Explaining Schizophrenia'). This point dovetails with the psychiatric emphasis on lack of insight and level of distress and dysfunction when diagnosing any major psychotic or mental disorder (APA 1994).

The second natural objection is that the clash between ways of knowing one's own mind is not present in all instances of alien thought: it is not necessary. This objection could concede that, as seems to be the case, the mental events disowned by schizophrenics tend to be persistent, intrusive, and antithetical to their avowed beliefs, desires, and emotions. But could there not be an alien thought that was not ego-dystonic in this way? Patient reports of ego-syntonic alien thoughts are extremely uncommon, but consider, for example, the following:

[S]he said that 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 it isn't the same ... the feeling is that it is somebody else's

⁷ A full account of the epistemology of self-knowledge will require an explanation of how a subject is able to distinguish between judgments that result from practical enquiry and those that express standing beliefs. Broadly speaking, there are three kinds of possible answer to this question. The first is purely phenomenological: one might suggest that there just is a salient conscious difference between them. The second is purely evidential: there is no reliable salient conscious difference, but we can use the context of a judgment, e.g., our memory of what we were doing before its occurrence, to infer whether it results from practical enquiry or expresses a standing belief. The third is a hybrid: both phenomenological and evidential considerations are relevant to how subjects make this distinction. I think it is likely that a good explanation will prove to be hybrid in form.

...’ She was asked if she had other people’s thoughts put inside her head. She said ‘ ... possibly they are but I don’t think of them in that way ... they were being put into me into my mind ... very similar to what I would be like normally’ (Hoerl 2001, 190).

But note two things about this report. First, the experience of disownership is being suggested to the patient by the interviewer. The patient is being encouraged to explore whether or not she experiences thoughts as not her own: she is being prompted. Second, the patient’s attitude to this suggestion is ambivalent if not indeed self-contradictory. She reports both that the thought does not seem like her own, and that it does. Indeed, the reason she suggests for it seeming like her own is precisely that it is not antithetical to what she herself thinks, but harmonious with it. Hence, in so far as there is evidence of ego-syntonic alien thoughts, it seems to point as much in favour of the account offered here as against it. Precisely because there is not an obvious clash, the schizophrenic is much less inclined to disown it. One final question that remains is why, if this account is correct, the schizophrenic privileges practical enquiry over introspection as a means of self-knowledge? Why not privilege introspection? Interestingly, Moran is adamant that practical enquiry does have a sort of priority; and that this is because of its importance to the psychic health of the person. But he does not offer a clear explanation of why this should be. However, I think it is possible, at this stage, at least to sketch an outline of an answer to this question. The answer is that practical enquiry is not just a source of self-knowledge, but a means by which one exercises rational control and thereby gains responsibility for one’s psychological life. It is therefore intimately linked to our potential for autonomy and freedom – both of which are, of course, compromised in individuals who are deeply mentally disturbed. In some sense, then, one might thus see the schizophrenic’s clinging to the priority of practical enquiry as a sign of hope and humanity: it reflects their potential for mental health.

5. The philosophy of psychiatry

The account of alien thought offered above accommodates the three, broad considerations that motivated its development. First, it allows for a unified account of the disownership of thoughts, impulses, and feelings. For all such mental phenomena can be manifestations or expressions of mental states which may lie outside of one’s rational control. Second, it opens up the possibility of explaining why a person who is schizophrenic disowns only a select array of mental events. As we saw, normally it is the case that the mental events treated as alien are antithetical to the schizophrenic: their content is in some way at odds with the schizophrenic’s self- or world- image or values, as expressed in their beliefs, desires, and emotions. It is therefore possible that psychodynamic exploration can shed light on why particular mental states which may be the cause of the alien mental events are, on the one hand present in the psyche, and, on the other, difficult to control through rational processes. Indeed, there is good evidence that talking therapies of various sorts have a positive effect on schizophrenic delusions (Dickerson 2000). This would be difficult, although not impossible, to understand if the explanation of schizophrenic delusions did not depend on personal-

level processes. Third, again as we saw, the proposal makes disownership an instance – no doubt extreme – of what is nonetheless an ordinary human experience: the failure of ideal rational control over one's psychological life. We already know that schizophrenics display exaggerated irrationality in belief formation: for instance, not only are they more likely to attribute causal responsibility to others, they are also more likely to jump to conclusions than normal subjects (Garety and Hemsley 1994). The proposal suggests that they are prone to alien thought because they may also show exaggerated irrationality in the capacity for conscious reflection to causally influence the maintenance and revision of beliefs and other mental states. That is something which in principle is open to empirical testing. But importantly, as I have emphasized, it is an abnormality that places them along a continuum together with the rest of us. Indeed, it is natural to envisage a spectrum of related abnormalities, more and less pathological, moving from immoral or selfish or shameful thoughts, to addiction and akrasia, to obsessive thinking and disorders, through to prodromal alien thought and finally full-blown schizophrenia.

Within the philosophy of psychiatry, and indeed psychiatry itself, there is a famous debate, inaugurated by Thomas Szasz, about whether or not mental illness is a myth (Szasz 1960; 1972). Many different ideas are captured by this slogan, and they are not always easy to separate (Pickard 2009). But one crucial thought contained in it is that our current diagnostic categories do not carve the world at its joints: they do not correspond to biologically definable diseases, but rather reflect the historical product of an attempt within psychiatry to collect together symptoms in a clinically – perhaps even politically – useful way. There are many different sources of evidence for this idea. One kind of source is scientific data. This includes data about (i) the lack of a single underlying genetic or neurological cause corresponding to disease categories; (ii) high levels of co-morbidity; (iii) the unreliability of diagnoses; (iv) discriminant function analyses that show that patients objectively cluster not according to diagnostic category, but rather according to type of symptom; (v) high levels of delusion in the general and prodromal populations, indicative of a continuum.⁸ Another kind of source is the history and sociology of psychiatry: the waxing and waning of diagnoses of particular diseases. Perhaps more than any other contemporary thinker, Ian Hacking has tried to make sense of this phenomenon. In his discussion of multiple personality disorder and fugue, Hacking very carefully tries to keep in sight the reality and horror of mental illness, while questioning the validity of psychiatric classifications by examining the history of particular diagnoses and changing conceptualizations of symptoms, together with how these changing conceptualizations allow the patient and the clinician to come to agree on a shared way of understanding and expressing what is wrong. To use his phrase, psychiatric categories are not natural but interactive kinds (Hacking 1995; 1998; 1999).

Alien thought is a diagnostically central symptom of schizophrenia. An anecdotal report: my uncle is a paranoid schizophrenic who has never suffered from alien thought; he has, however, consistently complained that his psychiatrists are always trying to get him to admit he suffers from alien thought. For it would certainly make his diagnosis much more certain and clear. In this respect, compare the ego-syntonic

⁸ See Bentall (2003) for a comprehensive survey.

final patient report, quoted above, where the interviewer is taking a strong lead. It is very important not to lose sight of the possibility that alien thought is a myth, in the sense that it does not correspond to a single, clearly circumscribed, psychological problem. Interestingly, one would never presume this sort of unity when considering other sorts of psychiatric problems, like, for instance the self-harming impulses and behaviour characteristic of Borderline Personality Disorder. On the one hand, it is an open question what exactly to count as self-harm. Cutting, burning, and other forms of clear and deliberate self-injury are paradigm instances of it, but arguably behaviour as diverse as physical self-neglect, substance abuse, recklessness and dangerous risk-taking, overeating, bingeing, and anorexia, should be included too. On the other hand, it is clear that different people self-harm because of different underlying causes and motivations, ranging across the gamut from stereotopic behaviour, to release, to distraction, to repression, to expression, to intentional communication. Similarly, it may be that alien thought collects together a gamut of disorienting and distressing ego-dystonic experiences – certainly it is not always clear how to distinguish some instances of alien thought from a supposedly different delusion, hearing voices – which, through interaction and dialogue with psychiatrists, patients come to call alien. The idea is that schizophrenics and psychiatrists together are using this as a label which they feel they mutually understand, precisely because it invokes the more ordinary sense of failing to endorse or identify with an aspect of one's mind.

It is certainly possible to envisage information-processing models of alien thought which allow there to be much in common between the schizophrenic and a more ordinary psychological life. But Frith's does not. That is something which has not, on the whole, been recognized. And it should be. For if Frith is right, then alien thought is wholly distinct from ordinary human experience, in much the way a cancerous body is distinct from health. There would be no continuum, in this respect, between mental disturbance and mental health. Correspondingly, there would be little scope for genuine empathy. This is unlikely to be true, and it should certainly not be presumed. One of the virtues of the alternative account I have offered is that it makes no such presumption.⁹

REFERENCES

- Aimola Davies, A. and Davies, M. 2009. Explaining pathologies of belief. In *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*, ed. M. Broome and L. Bortolotti, 285-323. Oxford: Oxford University Press.
- American Psychiatric Association (APA) 1994. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition. Washington DC: American Psychiatric Association.

⁹ Versions of this paper have been read in London, Bristol, and Oxford, and I am grateful to audiences for their interest and comments. I would also like especially to thank Tim Bayne, Lisa Bortolotti, John Campbell, Francesca Happe, Elijah Millgram, Luca Malatesti, Ian Phillips, Matt Soteriou, Fin Spicer, and Wayne Wu for extremely helpful criticism and discussion.

- Baker, C. A., and Morrison, A. P. 1998. Cognitive processes in auditory hallucinations: attributional biases and metacognition. *Psychological Medicine* 28(5): 1199-208.
- Bayne, T. and Pacherie, E. 2005. In defence of the doxastic conception of delusions. *Mind & Language* 20(2): 163-88.
- Bentall, R. 2003. *Madness Explained*. London: Penguin.
- Bortolotti, L. 2009. *Delusions and Other Irrational Beliefs*. Oxford: Oxford University Press.
- Bortolotti, L. and Broome, M. 2009. A role for ownership and authorship in the analysis of thought insertion. *Phenomenology and the Cognitive Sciences* 8(2): 205-24.
- Broome, M. and Bortolotti, L. (eds.) 2009. *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives*. Oxford: Oxford University Press.
- Campbell, J. 1999. Schizophrenia, the space of reasons, and thinking as a motor process. *The Monist* 82(4): 609-25.
- Davies, M. and Coltheart, M. 2000. Introduction: pathologies of belief. *Mind & Language* 15(1): 1-46.
- Davies, M., Coltheart, M., Langdon, R. and Breen, N. 2001. Monothematic delusions: towards a two-factor account. *Philosophy, Psychiatry, and Psychology* 8(2/3): 133-58.
- Dickerson, F. B. 2000. Cognitive behavioural psychotherapy for schizophrenia: a review of recent empirical studies. *Schizophrenia Research* 43(2/3): 71-90.
- Fernandez, J. 2010. Thought insertion and self-knowledge. *Mind & Language* 25(1): 66-88.
- Frankfurt, H. 1971. Freedom of the will and the concept of a person. *Journal of Philosophy* 68(1): 5-20.
- Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hove: Erlbaum (UK) Taylor & Francis.
- Gallagher, S. 2004. Neurocognitive models of schizophrenia: a neurophenomenological critique. *Psychopathology* 37(1): 8-19.
- Garety, P. A. and Hemsley, D. R. 1994. *Delusions: Investigations into the Psychology of Delusional Reasoning*. Oxford: Oxford University Press.
- Hacking, I. 1995. *Rewriting the Soul*. Cambridge: Harvard University Press.
- Hacking, I. 1998. *Mad Travellers*. Charlottesville: University of Virginia Press.
- Hacking, I. 1999. *The Social Construction of What?* Cambridge: Harvard University Press.
- Hoerl, C. 2001. On thought insertion. *Philosophy, Psychiatry and Psychology*, 8(2/3): 189-200.
- Marcel, A. 2003. The sense of agency: Awareness and ownership of action. In *Agency and Self-awareness*, ed. J. Roessler and N. Eilan, 48-93. Oxford: Clarendon Press.
- Mellor, C. H. 1970. First rank symptoms of schizophrenia. *British Journal of Psychiatry* 117(536): 15-23.
- Moran, R. 2001. *Authority and Estrangement*. Princeton: Princeton University Press.
- Moritz, S., Woodward, T. S., Burlon, M., Braus, D. F. and Andreson, B. 2007. Attributional style in schizophrenia: evidence of a decreased sense of self-causation in currently paranoid patients. *Cognitive Therapy and Research* 31(3): 371-83.
- Pickard, H. 2009. Mental illness is indeed a myth. In *Psychiatry as Cognitive Neuroscience*:

- Philosophical Perspectives*, ed. M. Broome and L. Bortolotti, 83-101. Oxford: Oxford University Press.
- Saks, E. 2007. *The Centre Cannot Hold: A Memoir of my Schizophrenia*. London: Virago.
- Soteriou, M. 2005. Mental action and the epistemology of mind. *Noûs* 39(1): 83-105.
- Sperry, R. W. 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology* 43(6): 482-9.
- Stephens, G. L. and Graham, G. 2003. *When Self-consciousness Breaks*. Cambridge: MIT Press.
- Szasz, T. 1960. The myth of mental illness. *American Psychologist* 15(2): 113-8.
- Szasz, T. 1974. *The Myth of Mental Illness*. London: Palladin.
- Von Holst, E. and Mittelstaedt, H. 1950. Das reafferenzprinzip. *Die Naturwissenschaften* 37(20): 464-76.
- Vosgerau, G. and Newen, A. 2007. Thoughts, motor actions, and the self. *Mind & Language* 22(1): 22-43.
- Wegner, D. M. 2002. *The Illusion of Conscious Will*. Cambridge: MIT Press.
- Wu, W. Under Review. Explaining Schizophrenia.

Received: January 28, 2010

Accepted: April 5, 2010

Oxford Centre for Neuroethics and
All Souls College
University of Oxford
OX1 4AL
United Kingdom
h.pickard@gmail.com