

University of Warwick institutional repository: <http://go.warwick.ac.uk/wrap>

This paper is made available online in accordance with publisher policies. Please scroll down to view the document itself. Please refer to the repository record for this item and our policy information available from the repository home page for further information.

To see the final version of this paper please visit the publisher's website. Access to the published version may require a subscription.

Author(s): Bortolotti, L. & M.R. Broome

Article Title: Delusional beliefs and reasoning giving

Year of publication: Not yet published

Link to published version: Not yet published

Publisher statement: None

Draft: Please do not quote without the authors' permission.

DELUSIONAL BELIEFS AND REASON GIVING

Lisa Bortolotti and Matthew Broome (forthcoming in *Philosophical Psychology*)

Abstract

Delusions are often regarded as irrational beliefs, but their irrationality is not sufficient to explain what is pathological about them. In this paper we ask whether deluded subjects have the capacity to support the content of their delusions with reasons, that is, whether they can author their delusional states. The hypothesis that delusions are characterised by a failure of authorship, which is a dimension of self knowledge, deserves to be empirically tested because (a) it has the potential to account for the distinction between endorsing a delusion and endorsing a framework belief; (b) it contributes to a philosophical analysis of the relationship between rationality and self knowledge; and (c) it informs diagnosis and therapy in clinical psychiatry. However, authorship cannot provide a demarcation criterion between delusions and other irrational belief states.

Draft: Please do not quote without the authors' permission.

Introduction

There is an extensive philosophical literature on the nature and rationality of delusions, aimed at establishing whether the behaviour of deluded subjects can be genuinely characterised in intentional terms and what norms of rationality deluded subjects violate (Stone & Young 1997; Campbell 1999; Davies & Coltheart 2002; Gerrans 2001; Bortolotti 2005; Broome 2004).

However, the fact that subjects with delusions fail to conform to basic norms of good reasoning is not by itself sufficient to explain the perceived difference between delusions and other irrational beliefs. The psychological literature on human reasoning has already shown that most if not all humans (a) fall prey to systematic mistakes in deductive, statistical and probabilistic reasoning (Stein 1996; Stanovich 1999; Samuels et al. 2002); (b) are vulnerable to reasoning biases; and (c) often exhibit inconsistencies between reported attitudes and between attitudes and behaviour (Thaler & Tversky 1990; Wilson & Hodges 1994). Although certain reasoning biases are more pronounced in deluded and high risk samples (Broome et al. 2007), the explanation of the puzzling nature of delusions has to be found elsewhere.

Subjects with delusions are found to form their delusional state on the basis of evidence that is not strong enough to justify its wildly implausible content. Moreover, they typically hang on to their beliefs tenaciously, in a way that is impervious to what others take to be powerful counterevidence or counterarguments. The extent to which this occurs is so significant that it has been hypothesised that they suffer from a specific

Draft: Please do not quote without the authors' permission.

deficit which prevents them from rejecting the delusional belief as false, once it is formed (Coltheart 2005). Finally, delusional states sometimes turn out to be badly integrated with other beliefs the subject has. However, even these failures at the level of epistemic rationality (Bermudez 2001), which manifest themselves in the processes of belief formation, revision and integration, are not different *in kind* from failures of rationality that all believers experience (Bortolotti 2002). So we need to look further for what is (qualitatively) distinctive about delusions.

The notion of authorship as the capacity to endorse a thought as one's own and justify it on the basis of reasons has been developed by Richard Moran and discussed widely in the philosophical and psychological literature in relation to the capacity for introspective reflection and for the capacity to exercise personal authority over reported attitudes (Moran 2001; Moran 2004; Lawlor 2003; Ferrero 2003; Carman 2003). This account of authorship, as we see it, is not necessarily tied to the rationality of the beliefs endorsed or to the rationality of the process of formation of such beliefs. Rather, authorship lies in the capacity to endorse the content of a belief and justify it with reasons. Moran's thesis is that authorship generates a form of first-person authority. By taking responsibility for the thoughts one reports, one relates to the content of those thoughts in a way no other person can. Due to the identified link between reason giving and self knowledge, an examination of the effects of reflection on the consistency of reported attitudes and behaviour has already stirred an interesting debate on the limitations of self knowledge in the context of the introspection studies.

Draft: Please do not quote without the authors' permission.

We argue that importing the notion of authorship into the study of delusions makes the following contributions to our understanding of pathological beliefs. First, an empirically informed philosophical investigation will help establish whether there is a connection between rationality and self knowledge, on the basis of the hypothesis that failures in reason giving impair both capacities. One issue to explore is whether subjects maintain personal authority over a belief that they are virtually unable to reject and, further, whether they can make autonomous decisions on the basis of that belief.

Second, given the implications of authorship for autonomous decision-making and action, considerations about whether subjects with delusions are authoritative with respect to their delusional beliefs can be used to assess and inform current methods of diagnosis and give rise to a review of the appropriateness of current therapeutic interventions.

Finally, considerations about authorship shed some light on the thorny classificatory issues surrounding the notion of delusion, including the question whether all reported delusional states deserve intentional characterisation. A related question is whether failures in reason giving are a promising candidate for a demarcation criterion between delusional and non-delusional beliefs. On the basis of the examples that will be considered in this paper, we shall conclude that delusions differ greatly in their authorship-potential, which means that identifying failures in reason giving will not be sufficient to distinguish delusions from other irrational beliefs.

The analysis we provide in terms of authorship confirms that there is a range of potential failures of rationality and self knowledge in the reporting and endorsement of

Draft: Please do not quote without the authors' permission.

pathological beliefs, but no unique demarcation criterion can be identified. Some subjects of delusions maintain the capacity to argue for the content of their delusional states and give reasons for endorsing it. Other subjects with delusions experience a radical breakdown of self knowledge which prevents the intentional characterisation of their delusional reports. If this variation is observed, neither irrationality nor lack of authorship seems to be a promising candidate for a demarcation criterion between delusions and other beliefs. It would not be surprising if the project of identifying a single defining characteristic of delusional beliefs turned out to be misguided, as there is continuity between pathological and other irrational beliefs.

One caveat is in order. Many have argued that delusions can never be coherently described as belief states (Berrios 1991); that they constitute a propositional attitude significantly distinct from that of ordinary beliefs (Stephens & Graham 2006); or that they demand a metacognitive explanation (Currie 2000). For those who are convinced by one of these accounts or by any other non-doxastic account of delusions, our attempt to import a notion that typically applies to beliefs, that of authorship, into the study of delusions seems to involve an unmotivated assumption. But our claim that delusions are *prima facie* belief-like states is not an assumption and is not unmotivated. We argued for the continuity between belief states and delusional states elsewhere (Broome 2004; Bortolotti 2002; Bortolotti 2005; Bortolotti & Broome 2007) and we follow the defenders of the two-factor account of delusions in considering this continuity as the most promising starting point for a satisfactory explanation of delusions (Bayne & Pacherie

Draft: Please do not quote without the authors' permission.

2005; Davies et al. 2002; Coltheart 2005). It is worth noticing that, in some of the accounts that put pressure on the continuity between ordinary beliefs and delusions, the notion of belief deployed is often extremely narrow or extremely demanding, making the dispute more of a terminological than a substantive one. For instance, some authors who argue that delusions cannot be beliefs because they do not guide action, neglect to acknowledge that many ordinary beliefs have a very tenuous action-guiding role and talk as if delusions were the sole manifestation of psychopathology in schizophrenia. By not paying attention to the other manifestations of schizophrenia, they fail to acknowledge the behavioural effects of deficits in affect and motivation which can be present independent of delusions and can be the cause of delusions not being acted upon (Stephens & Graham 2004; Berrios 1991; Sass 1994). However, it is possible that those subjects who do not act on their delusional beliefs, or those who view their delusions ironically or dispassionately, do so as a result of the other deficits in the syndrome, and not as a result of their failure to genuinely believe.

In this paper, we shall explore an aspect of the pathology of belief that is viewed as a breakdown of self knowledge and see whether it deepens our understanding of what is amiss in many delusional reports, with a view to stimulating further empirical research. Although the analysis will have greater appeal for those who endorse a doxastic account of delusions, it should be of interest also for their opponents, for the following two reasons. First, an analysis of delusions in terms of authorship acquires credibility in a context in which it is acknowledged that at least some delusions behave like belief states,

Draft: Please do not quote without the authors' permission.

but in the end makes intentional characterisation hostage to the degree of authorship subjects manifest towards their delusional beliefs and thereby suggests that some of the reports that are accompanied by a radical breakdown of self knowledge cannot be ascribed to the subjects as beliefs. Second, authorship is not exclusively a feature of beliefs (as we shall see) and therefore can be a useful explanatory tool even for those philosophers and cognitive psychologists who are more sceptical about the continuity between delusions and ordinary beliefs.

1. Authorship

1.1. Which states can be authored and how?

Andrew's story about why he applied to study law at university is different from the way in which his friends explain his decision. He justifies his decision as a good move for his future career, whereas his friends suspect that he chose to study law at his father's insistence. First-person and third-person accounts of behaviour diverge for at least two important reasons. Andrew knows things about what he thinks and how he feels that others can only derive by observing his behaviour and making inferences from it. He has some prima facie epistemic advantage over his friends, given by direct access to the content of some of his mental states. Moreover, Andrew is vulnerable to different attribution and reasoning biases from the ones that affect third-person observers. For instance, he will be exposed to self attribution biases when he reflects upon his own

Draft: Please do not quote without the authors' permission.

dispositions, personality traits, past history, likelihood of success, and so on. Hence, access to the content of his mental states is not entirely transparent, but rather, his knowledge of his own reasons is refracted through such biases. It is not easy to determine which story is the correct one if conflict between first- and third-person accounts emerges, and it is equally hard to come up with criteria of correctness for the ascription of mental states to people, given that psychological explanations for action are complex and that people's mental life is not static. On the contrary, dispositions, traits and even biases are subject to change.

The notion of authorship as developed by Moran suggests that there is something special about Andrew's own explanation for his decision. Andrew takes an agential perspective on his decision by endorsing it as the right decision, whereas his friends can only be observers of his behaviour and interpreters of the reasons for his decision. Whatever story he tells about it, Andrew is 'authoritative' about his decision, because he can do something that his friends cannot do; endorse the decision on the basis of his sense of his best reasons for it. In other words, by being able to offer a justification for it, Andrew is in a position to take responsibility for his decision.

Moran's message is that self knowledge is not exhausted by direct epistemic access to one's thoughts and self interpretation of behaviour on the basis of evidence. Self insight into mental states, such as beliefs and preferences that are central to our concerns and interests, is questioned when the content of these states is neither endorsed nor defended

Draft: Please do not quote without the authors' permission.

with reasons. Reason giving offers one an additional route to knowing the content of the mental states one endorses.

If it is possible for a person to answer a deliberative question about his belief at all, this involves assuming an authority over, and a responsibility for, what his belief actually is. Thus a person able to exercise this capacity is in a position to declare what his belief is by reflection on the reasons in favor of that belief, rather than by examination of the psychological evidence. In this way [...] avowal can be seen as an expression of genuine self knowledge. (Moran 2004, 425)

But authorship does not apply to all the mental states one is aware of. Some mental states are not candidates for authorship or can be authored only to a very limited extent, as one does not typically deliberate upon or justify the content of a merely perceptual belief or a drive. (Of course, there are exceptions. In the context of a philosophy class on scepticism or a conversation with a solipsist, defending the content of one's perceptual experiences with reasons is required.)

Andrew would not have self knowledge with respect to his belief that studying law is the best option for him, unless he were in a position to defend that claim with reasons. But the same would not apply to other conscious mental states of Andrew. His being hungry or his seeing his bike leaning on the front gate of the school are states that he can authoritatively report without giving reason for them; it is not so clear what it would mean to 'endorse' them, other than acting in a way that is consistent with them.

There are also less perceptually-driven beliefs one has without being able to offer independent reasons for having them (in the literature they are referred to as 'framework

Draft: Please do not quote without the authors' permission.

beliefs' or 'hinges'). One must believe that the earth existed prior to one's birth, for instance. Some philosophers have discussed whether agents who are typically very committed to such beliefs are in a position to engage in the game of justification with respect to them (Wittgenstein 1969; Campbell 2001; Eilan 2000; Thornton forthcoming; Bortolotti 2002). Any reason speakers could offer to justify the believed proposition that the earth existed before their births is reliant upon the same proposition being true.

Thus, the acceptance of framework propositions is usually not questioned and is considered to be immune from doubt. Can one author a framework proposition? There are conflicting answers in the literature. Let's think about what it would be like to justify the belief in a framework proposition. The belief that the earth existed before one's births can be defended only on the basis of coherence with other beliefs one has and not on the basis of there being more basic reasons to support the conviction that the proposition is true. But a defence based on epistemic conservatism or coherence is still a form of justification.

One is usually not asked to justify framework beliefs, but if one were, one could do so. That said, the reasons adduced in support of the beliefs would not probably persuade the sceptic of the truth of the framework propositions. This tells us something interesting about authorship in general: it is not about grounding conceptually sophisticated beliefs onto more basic ones or answering sceptical challenges, but is primarily about endorsement. That is why the nature of the endorsement of framework propositions is also relevant in this context. One can manifest the endorsement of the claim that the earth

Draft: Please do not quote without the authors' permission.

existed long before one's birth, both in behaviour and in reporting other beliefs grounded on that firm conviction, as for instance the belief that one's parents were born before one's birth, even without ever explicitly reporting that belief. A sign of this tacit endorsement is the use of a shared language that presupposes the truth of the relevant framework proposition. Many terms in one's language, including the terms 'parent' and 'history', would acquire a different meaning if the framework proposition turned out to be false.

The account of hinges as framework propositions has been challenged in recent work. According to Moyal-Sharrock (2004), the endorsement of a hinge is deemed as necessarily ineffable. If the endorsement of hinges cannot be reason-based and cannot be articulated or verbalised, then it may not allow for authorship, as the type of endorsement that usually characterises authorship involves the capacity for justification via reason giving. On this alternative account, the endorsement of a hinge would be manifest primarily in a way of life, via non-verbal behaviour, rather than via the capacity for arguing for the content of a belief. As it is not clear whether hinges could be object of belief, the potential for authoring them is compromised in this case.

To sum up what we have said so far about authorship: (1) authorship of a belief is achieved via reason giving, that is, a subject is the author of a belief if she is able to provide reasons for endorsing that belief, either in deliberation or justification; (2) authorship typically applies to conscious mental states on which content one can exercise some control, such as beliefs, preferences and decisions, but some mental states can be

Draft: Please do not quote without the authors' permission.

authored to a greater degree than others; (3) although reason giving is necessary for authorship, justification does not need to meet foundationalist standards; (4) the endorsement of a belief which results from authorship can be explicitly or implicitly manifested in several ways, e.g. behaving in a way that is compatible with the endorsed beliefs.

In our application of the notion of authorship to some cases of delusions, we shall come back to these four points and to whether both the level of commitment to one's beliefs exercised via reason giving and the manifestation of one's commitment vary according to the type of belief one reports. But first, we need to be explicit about the conditions for authoring beliefs.

1.2. What are the conditions for authorship?

So far we have been presenting a notion of authorship that is largely borrowed from the work of Moran. We shall now take some liberties with the notion of authorship developed by Moran, by revisiting the controversial issue of the relationship between authorship and rationality. In what sense, if any, is rationality constitutive of authorship?

Moran's account of authorship is interpreted as strongly idealised, as he characterises deliberation and justification, the two processes by which beliefs are authored, as the exercise of reason. But it is in the spirit of how Moran argues for the pervasiveness of the

Draft: Please do not quote without the authors' permission.

notion of authorship to claim that, in order to be the author of a belief, all the rationality or reasonableness required is that an agent takes herself to have better reasons for endorsing that belief than not. It is neither necessary that the belief be optimally rational nor that the processes of deliberation and justification which lead one to endorse the belief be themselves optimally rational.

Now let's consider two versions of the conditions for authorship as applied to our previous example:

- a. In order to be the author of the belief that he should study law at university, Andrew needs to be able to endorse the content of his belief on the basis of his best reasons for it.
- b. In order to be the author of the belief that he should study law at university, Andrew needs to be able to endorse the content of his belief on the basis of what he takes to be his best reasons for it.

The formulation in (b) differs from that in (a) in the way in which rationality is seen as a precondition for authorship. There are a number of relevant questions about the rationality of an act of authorship: whether the authored belief is rational; whether the process of deliberation or justification leading to the endorsement of that belief is rational; or whether the reasons for endorsing the belief track the best available evidence. We shall focus here on the third sense of rationality.

One should endorse a belief on the basis of the best available evidence, but *available to whom?* To an impartial observer in ideal epistemic conditions or to oneself, that is, the

Draft: Please do not quote without the authors' permission.

person who is aspiring to become the author of the belief? One can be mistaken about which reasons are the best reasons for endorsing a given belief, but as long as one takes oneself as having the best reasons for the belief, one engages in a genuine act of deliberation or justification.

As we want to be able to account for authorship as a notion that admits of degrees, then (b) still demands too much as a formulation of the conditions for authorship. One may have to make a commitment to the truth of a statement in absence of relevant information or decide on a course of action in conditions of partial ignorance about present or future events. Even in the lights of the person making the commitment, the reasons appealed to in the act of deliberation or justification might not be very strong reasons, and yet the commitment or decision needs to be made. To account for these frequent situations, we propose to weaken the rationality condition for authorship in the following way:

c. In order to be the author of the belief that he should study law at university, Andrew needs to be able to endorse the content of his belief for reasons that he takes to be better than either his reasons for the belief that he shouldn't study law, or his reasons for suspending his judgement about it.

Now that the conditions for authorship have been spelled out more clearly, we can see how the notion of authorship can apply to the reporting of delusional beliefs.

Draft: Please do not quote without the authors' permission.

2. Reason giving in delusions

There are a variety of ways in which delusions are classified in psychiatry. Jaspers' and subsequent accounts rely on the distinction between 'form' and 'content' to guide classification. The content is thought of as the theme of the delusion, such as persecution, control, infatuation. The form concerns the structure of the belief and its relationship to the reasons the subject can offer the interviewer for holding it. In particular, form is affected by whether such reasons are 'understandable' or not. When these reasons are deemed 'un-understandable' the delusion is said to be primary or autochthonous. Hence, form has been consistently viewed as of particular importance both in terms of diagnosis but also prognosis. Content has been seen as somewhat more 'epiphenomenal' and related to the subject's biography, concerns and culture, whereas the form has been regarded as reflecting pathological processes.

This type of classification is promising for our purposes as it seems to track a notion, that of understandability of the delusional belief, that easily correlates with the capacity a subject has to present reasons in support of the reported delusional belief. The more a subject is able to support the content of her delusion with reasons, and to manifest this endorsement by integrating the delusional belief in a narrative about herself that seems coherent to others, the more the delusion is understandable and amenable to being characterised in intentional terms by an interpreter.

This point suggests that rationality, self knowledge and even intentionality are interconnected: the type of commitment made to a reported belief-like state raises issues

Draft: Please do not quote without the authors' permission.

about the appropriateness of intentional characterisation of that state. Moreover, its manifestations in terms of the integration of the belief-like state in a system and its inferential relations with other belief states contributes to assessing the subject's behaviour in terms of coherence and rationality. Even more important for our purposes, commitment to a reported belief-like state via reason giving offers an additional first-personal route to the knowledge of the content of that state and to the knowledge that one is having that state, thereby giving rise to first-person authority and affecting judgements of self knowledge.

We propose that the quality of the endorsement of a belief-content via reason giving helps establish (i) whether the belief is genuine; (ii) whether it has been rationally formed, it is integrated in an existing system of beliefs, and it has the appropriate action-guiding features; and (iii) to what extent the subject has first-person authority over it. We shall proceed by considering the form of some reports by people affected by delusions and ask to what extent these people are committed to the content of their delusion.

Before proceeding, two methodological caveats are in order. First, the list of cases reported has no pretence to be exhaustive and the cases chosen are not to be regarded as especially representative of a typology of delusion. They have been chosen because they cover a range of delusional reports which differ in terms of the contents of the delusional states; their significance for the subjects; and the degree of elaboration manifested.

Second, any conclusive remark about the role of authorship in the classification of delusions or about the potential for our analysis in terms of authorship to inform

Draft: Please do not quote without the authors' permission.

diagnosis and therapy will have to be supported by specifically tailored empirical data about the way in which people defend the content of their delusions when prompted to give reasons for them. As it stands, this is a speculative exercise which fulfils two purposes: (a) it offers a *prima facie* argument for the potential benefits of introducing the notion of authorship in the delusion debate; and (b) it encourages researchers in this area to extend empirical investigations of reason giving, of the type that has been conducted with respect to reported attitudes in normal psychology in the introspection studies, to the domain of pathological beliefs.

2.1. Justifying delusional states

Consider the following textbook examples of delusions:

(a) A 21-year old man has sudden conviction that certain songs played on the radio used his voice in the role of lead singer. He cannot explain why. (Yager & Gitlin 2005, page 978)

(b) A man believes his wife is unfaithful to him because the fifth lamp-post along on the left is unlit. (Sims 2003, page 119)

(c) A woman claims her blood is being injected out of her body in her sleep because she has spots on her arms. The interviewer says that they are freckles and that he has them

Draft: Please do not quote without the authors' permission.

too – and shows her. She agrees they are similar, but still maintains that she is being injected. (Sims 2003, page 123)

(d) A female subject does not recognise the man who drives the family car and pays the bills as her husband, although she concedes that he looks very much like her husband. She notices that he is just a bit 'fatter'. (Sims 2003, page 123)

In cases such as (a) and (b) subjects make the reports with conviction but offer no explanation or reason to believe that what they say is true. The inability or unwillingness to provide evidence for what they claim is somehow at odds with the subjects' apparent confidence and with the personal references contained in their reports.

Cases (a) and (b) illustrate delusions that, in Jaspers' terminology, are deemed 'ununderstandable', when no further elaboration can be offered. Cases (c) and (d) are different in so far as subjects make a report with conviction and are also able to defend the content of the beliefs they report by adducing facts that have a meaningful connection to them. The way in which reports differ from ordinary belief reports is that the reasons to which subjects appeal are not, all things considered, likely to be regarded by others as good or sufficient reasons to justify or maintain beliefs with that content.

It is very difficult to draw a sharp distinction between delusions on the basis of whether their content defies understanding, as cultural and individual differences can affect the plausibility of the delusional content itself or the reasons for it. But these difficulties impact on our analysis to a lesser extent if we focus not just on plausibility of content, but

Draft: Please do not quote without the authors' permission.

especially on the extent to which subjects with delusions are prepared to defend the content of their delusion. What seems to be at issue is not whether others are willing to accept the delusional belief as true when it is reported, but whether the very subject reporting it thoroughly *accepts* the belief. The degree of endorsement or acceptance is difficult to establish and only further examination of a subject's set of beliefs and carefully directed questioning can tell us whether the reasons adduced for endorsing the content of the delusional state are (subjectively) good reasons. Whether a reason is good in this sense depends on the number of meaningful connections that can be identified between the content of the delusion and other intentional states the subject is committed to.

For instance, for each of the cases in (a) and (b), there are two possible scenarios to be considered. In the former, the man who believes that his wife is unfaithful (hypothesis) because the lamp-post is unlit (evidence) refuses to give any further explanation of the relation between the two. In the latter, he reports the same belief but also elaborates a complex theory about the connection between the hypothesis and the evidence for it. Typical explanations could invoke intervention from a third party (e.g. God or the government is trying to warn him of his wife's infidelity by leaving clues that only he can decipher, as that particular lamp-post being unlit). In this latter case, the delusion appears as more understandable to an interpreter, because some meaningful connections between hypothesis and evidence are established, even if the plausibility of the hypothesis given the evidence is highly questionable. In this latter case, the interpreter gets a better sense

Draft: Please do not quote without the authors' permission.

that the speaker really believes the content of the reported delusion, and that he has been thinking about the connection between hypothesis and evidence (even if just at the stage of *post-hoc* rationalisation).

There are further obstacles that might hinder the attempt to distinguish cases in which no reasons for a belief are offered and cases in which the reasons offered are (subjectively) very bad reasons. Some would claim that a very bad reason for a belief ceases to be a reason for the belief altogether. The difference between the example (d) and (b) in this respect seems striking: the fact that an observed man is fatter than her husband can be conceivably regarded by a woman as a reason for denying that the man is her husband, whereas the lamp-post being unlit would seem to bear no meaningful relation whatsoever with whether a spouse is faithful, if no additional explanation is offered.

Case (c) sits somewhere between (d) and (b). There are recognisable meaningful links between blood being injected out of a woman's body and this woman having spots on her arms, but the type of spot identified as evidence for the injections is not of the right kind. As we were very explicit that rationality as optimisation (i.e. having the best available reasons for a belief, or even just objectively good reasons) is not a requirement for authorship, delusions in (c) and (d) could be regarded as authored if an attempt to engage in reason giving is made.

What about delusions in (a) and (b)? In these two cases the delusion is not authored, if no justification is offered for the content of the delusion. And yet the type of belief involved would demand justification and an expression of commitment, because its content has

Draft: Please do not quote without the authors' permission.

potentially very significant or disturbing implications for the subject (and the subject's conception of herself or her own life).

But one could object to this analysis on the ground that some types of beliefs, even when important in the economy of one's belief system, do not require justification and the fact that they are not authored should not be seen as a failure or a shortcoming. We did concede previously that issues about authorship are muddled in those circumstances in which we are so certain about the truth of the believed proposition that that in no circumstances we would be prepared to give it up, because in those cases there is no explicit engagement in reason giving. Couldn't delusions be a bit like framework beliefs?

2.2. Delusions and framework beliefs

Some attempts at explaining what causes failures in reason giving invite us to draw an analogy between beliefs in propositions such as 'The earth existed long before our births' and delusions. A subject reports her delusion with great conviction but refuses to offer reasons for it, because she believes that the fact reported is more certain than any of the facts that might lend support to it. Not to believe the content of the delusion is, for the subject, out of the question. Here the analogy lies in the fact that neither the delusion nor the framework belief is justified because there is a sense in which they cannot receive any independent justification.

Draft: Please do not quote without the authors' permission.

John Campbell makes this point in a recent and much discussed paper. He defends the analogy on the basis of the certainty with which the relevant proposition is believed:

The kind of status that we ordinarily assign to propositions like 'The world has existed for quite a long time' ... is assigned by the deluded subject to propositions like 'I am dead' or 'My neighbour has been replaced by an impostor.' That is, they are treated as the background assumptions needed for there to be any testing of the correctness of propositions at all. (Campbell 2001, page 96).

Here is another element of continuity between framework propositions and delusions.

The subject offers reasons that appear to have no meaningful connection to the content of the delusion, because the connection she identifies is based on a meaning relation that is salient to herself only and, as a result, the endorsement of the delusion is not understood by others. Here the analogy lies in the fact that the rejection or acquisition of a framework proposition has the same effect as the rejection or acquisition of a delusional belief. They both cause the subject to review meaning connections between previously used terms. For instance, we could no longer make sense of what the words 'parents' or 'history' meant if we did no longer assume that the earth existed before we were born.

It is certainly true that the content of some delusional beliefs causes a redefinition of meaning relations and implies an attribution of special significance to events might appear irrelevant or casual to anybody but the subject of the delusion. The lamp-post being lit *means* infidelity. Having freckles in one's arms *means* that injections were made. Or, as in Campbell's example, 'death' no longer *means* death.

Draft: Please do not quote without the authors' permission.

Sometimes the patient will say that she is disembodied. The patient may claim to be dead despite being able to walk and talk. Indeed, the patient may say that she is dead even though she realizes that no one else would accept this claim. The trouble is, how can the patients really be said to be holding on to knowledge of the meanings of their remarks when they are using the words in such a deviant way? (Campbell 2001, page 91)

The analogy between delusions and framework beliefs appears to be working and to support the view that, whatever we say about the authorship of framework beliefs, we should also say about the authorship of those delusions which share these fundamental features with them. But the analogy is far from water-tight at a closer look.

First, the refusal or incapacity to provide a justification for the delusional beliefs on the basis of their being absolutely and undoubtedly true is not sufficient reason to claim that the delusional beliefs play the role of background assumptions in reason giving – that would seem like a very implausible claim especially where delusions are relatively circumscribed (Broome 2004). We would expect a framework proposition to be perfectly well-integrated in the belief system. Indeed, the framework proposition should be one of its unshakeable pillars, and to be relied upon in the justification of beliefs that are much more open to doubt and to revision. But many primary delusions do not behave that way. They are isolated from the other beliefs in the system and no reason giving exercise employs them as assumptions.

Second, there is a problem with the analogy when we consider different types of delusions. Bayne and Pacherie (2004) have argued convincingly that, in some versions of

Draft: Please do not quote without the authors' permission.

the Capgras syndrome, the example originally used by Campbell in the first quotation we reported, delusions cannot be regarded as playing the same role of framework beliefs. Subjects with Capgras often recognise that the content of their delusion is implausible, sometimes even unbelievable, but they would not find the content of framework beliefs implausible or unbelievable.

Third, Campbell seems right that the adoption of a delusional belief leads subjects to subscribe to new meaning relations or reject previously accepted ones with respect to the Cotard delusion, where the subject claims to be dead. But the same phenomenon is totally absent in other delusions such as Capgras. As Broome (2004) argued, the explanation of what happens to a subject with Capgras, such as in example (d) above, does not usually need to rely on the introduction of new meaning relations. It is because subjects understand what the word 'impostors' means that they find it disturbing that their dear ones have been replaced by such impostors. It is because psychiatrists understand what subjects report that they are able to identify the delusion as a version of Capgras and ask more about it.

These objections to the analogy between delusions and framework beliefs can be reformulated and made better sense of by using the notion of authorship. The analogy fails because the nature of the commitment to a framework proposition is different from the nature of the commitment to the content of a delusion, and this seems to be true across the board, independent of what type of delusion we are considering. In what follows we shall pay attention both to issues concerning the degree of commitment as

Draft: Please do not quote without the authors' permission.

measured by the capacity for reason giving and the manifestations of that commitment in terms of other verbal reports and subjects' behaviour in general.

Let's focus on the intersubjective dimension of the relation between the subject and the proposition and consider both the issue of the redefinition of meaning relations and that of the verbalisation and communication of reasons. Changes in the commitment to delusions and framework propositions give rise to changes in meaning relations, but for very different reasons. Why do we think that giving up a framework proposition would lead us to redefine meaning relations? Because, just like language, the commitment to framework propositions is usually shared by an entire community – sometimes as large as that of all humans. Campbell (2001) offers the same answer to this question: abandoning a framework belief makes a difference to other beliefs and causes a redefinition of relevant terms in the language of the community that undergoes this often radical change. Philosophers often express the idea that meanings are shared by using slogans, such as 'meanings ain't in the head', or 'there cannot be a private language'. But there is a sense in which meaning relations revealed in the reports of delusions can be private to the subject with the delusion. The delusional belief is typically rejected by the people surrounding the subject, and the subject is acutely aware of that. This is a first important disanalogy between framework beliefs and delusions: they both bring about changes in meaning relations, but the former do so within a community whereas the latter do so for a subject in isolation from, and often opposition to, the community of which the subject is a member (Thornton forthcoming).

Draft: Please do not quote without the authors' permission.

There is another apparent analogy between framework beliefs and delusions that cannot resist further scrutiny: both are belief-like states for which commitment is difficult to properly articulate or verbalise. The subject's endorsement of a delusion can be difficult to verbalise or articulate clearly and therefore it is seldom communicated to others.

Indeed, these are some of the causes of distress and social alienation that often accompany the experience of subjects with delusions. A framework belief is almost never explicitly reported within a community and opportunities for expressing a commitment to it are rare. But again the reasons why the commitment to the belief-like state is not explicit are radically different in the two cases: the subject reporting the delusion often lacks the capacity or the will to share with others a belief that is likely to encounter incredulity; the subject endorsing a framework belief sees no need to express her commitment to it.

The former point concerns the manifestations of authorship. It is a consequence of the act of endorsement via reason giving that the belief endorsed is well-integrated in a system of beliefs and its inferential relations with other beliefs are invoked and worked out. This leads in some cases to changing meaning relations that depend on the commitment to the content of that belief (see the Cotard delusion and the meaning of 'death'). In other cases, there is no evident change in meaning relations, and the analogy completely breaks down. It is important to notice that, even when delusions cause alienation, meanings are not always private to the subject (see the Capgras delusion and the meaning of 'impostor'). The notion of a 'hinge' and of a 'framework' might not be that useful in relation to these

Draft: Please do not quote without the authors' permission.

types of delusions because their content is reported and communicated successfully to others, even if not endorsed by others. Meaning has to be shared to some extent: as the psychiatric adage goes, 'patients become well enough to tell us how sick they are'. For psychosis more broadly the thought that meaning relations are completely redefined and communication is impaired is more plausible. Arguably, in the most acute stages of mental illness, characterised by marked thought disorder, bizarre behaviour, and catatonia, the form of life that is occupied by the subject is incommunicable to others.

The latter point concerns the very definition of authorship as commitment achieved via reason giving. The subject affected by delusions often is unable to provide reasons for the reported belief. The subject who doesn't explicitly endorse a framework proposition is able to do so on request, but finds the articulation of reasons for the belief unnecessary. The commitment to a framework proposition is not always explicit and verbalised, but, when it is expressed in behaviour and in the endorsement of beliefs that rest on it, it brings social inclusion and participation in social practices. Endorsing the content of a delusion seems to have the opposite effect. It brings social exclusion and initiates a renegotiation of meanings that leads to a drastic revision of the belief system when the delusion is elaborated, or a drastic compartmentalisation when the delusion is circumscribed. In short, in the case of delusions regarded as ununderstandable either the reason-giving exercise required to author a delusion is not present, or the endorsement of the belief is 'private', that is, deprived of its social dimension. When reasons are given, they are not recognised as reasons by others.

Draft: Please do not quote without the authors' permission.

Let's move to considering the intrasubjective character of the commitment to a belief and concentrate once more on the manifestations of authorship as reason giving. Within the subject holding onto a belief, there are significant differences between the attitude towards a framework belief and the attitude towards a delusion. The commitment to framework proposition is pervasive and manifested in many instances of behaviour, although the belief remains in the background and may never be explicitly reported or justified. Not only is the believed proposition compatible with other propositions believed by the subject, but it brings cohesion within the system by supporting them and making their acceptance possible. In other words, the framework beliefs are used in reason giving when the commitment to other beliefs is tested.

Differences emerge between subjects with elaborated or circumscribed delusions. It is possible that to some extent subjects with elaborated delusions behave towards them as ordinary believers behave towards framework beliefs, by taking them for granted, employing them in reason giving and reserving to them a pronounced action-guiding role. But for the subjects affected by those delusions that are heavily circumscribed, the delusion does not give rise to the behaviour that would naturally ensue from believing with conviction the content of the delusion and the endorsement of the delusion is not implicit in the endorsement of other beliefs. Rather, the delusion often explicitly conflicts with other accepted beliefs, generating tensions and sometimes even inconsistencies in the belief system.

Draft: Please do not quote without the authors' permission.

To bring together inter- and intrasubjective considerations, the analogy between framework beliefs and heavily circumscribed delusions works if confined to observations about meaning change and gaps in the verbalised attempts to support the belief, but it fails to account for the intersubjective dimension that the acceptance of a framework proposition often has. The analogy between framework beliefs and more elaborated delusions makes sense of the continuity in the behavioural manifestations of the commitment to the belief at the intrasubjective level, but it is implausible at the intersubjective level where the changes of meaning relations required to make sense of the delusion are not shared by the rest of the subject's community and are not as easily communicated to others.

Rejecting the analogy, at least in some cases, has important consequences for authorship in delusions. There is a sense in which people who endorse a framework belief are 'excused from' defending it via explicit reason giving, due to the foundational nature of the belief. These subjects are still able to manifest behaviourally their commitment to the belief, and, in those situations in which giving reasons were acceptable, because of a temporary suspension of the special status of the belief, they would be able to come up with reasons. But it is important to notice that even framework beliefs can be subject to revision or object of dispute. For instance, the existence of physical objects is put to the test in a class on scepticism and the age of the Earth is debated between evolutionary biologists and creationists.

Draft: Please do not quote without the authors' permission.

No such excuse applies to delusions, because delusions do not play the foundational role of framework beliefs. Hence, the failures of reason giving that we often encounter in reports of delusions do demand an explanation. In the four cases above endorsement is present or manifested to different degrees. In reports of delusions such as (a) and (b), when no additional explanation is provided, there is no evidence of any capacity for reasons-based endorsement. In cases such as (c) and (d) subjects do seem to be able to give reasons. Similarly, in cases (a) and (b) there are no clear manifestations of endorsement, whereas in cases (c) and (d) subjects take some responsibility for the content of the belief that they report. Unfortunately no hypothesis can be made about the behavioural implications of the reports due to lack of details on the individual cases.

3. Potential implications

Why does it matter whether subjects author their delusional beliefs? There are several significant consequences of a failure in the capacity of giving reasons in support of one's belief.

One link that philosophers have already started to explore is the relation between endorsing a belief on the basis of reasons and having first-person authority over that belief. If it is true that, as Moran says, authorship offers an additional route to the knowledge of the content of one's beliefs, based on the evidence one has for endorsing those beliefs, then a failure of authorship would cause a breakdown of self knowledge.

Draft: Please do not quote without the authors' permission.

The hypothesis that subjects with delusions sometimes experience a breakdown of self knowledge fits very well with some behavioural manifestations in psychopathology, especially explicit inconsistencies in the beliefs reported and missing links between some beliefs and action. When subjects make conflicting claims in the course of the same interview, or are convinced that the person living with them is an impostor, and not their spouse, but do not make any attempt to look for their spouse elsewhere, they are not completely aware of the import of their delusional beliefs. In a recent paper, Coltheart (2005) argues that delusions are characterised by the fact that subjects cannot reject them: delusions become invulnerable and this can explain why they resist counterevidence and counterargument. But again, issues of self knowledge and rationality emerge in this context: can one say to be authoritative about one's own thought, if the thought becomes impervious to any consideration against it? It seems as if the thought is almost imposed on the subject, rather than owned by the subject, and this does have far-reaching implications for the notion of autonomy as applied to those beliefs and intentions that the subject forms on the basis of the delusional belief.

These issues are of great theoretical importance, as studying the relation between rationality and self knowledge in pathological cases can enlighten us on the mechanisms that support ordinary cognition. But they have a practical dimension too. Diagnosis and therapy in psychopathology would benefit from a closer examination of reason giving in delusional reports. Contemporary models of the pathophysiology of psychosis (Broome et al. 2005a; Garety et al. 2007; Kapur 2003; van der Gaag 2006) suggest that dopamine

Draft: Please do not quote without the authors' permission.

may play a central role in the inappropriate attribution of salience to representations. It has been argued that in the normal individual, mesolimbic dopamine acts to provide significance or salience, transforming an affectively neutral mental representation of a stimulus into an attractive or aversive one (Berridge & Robinson 1998). If psychosis were associated with increased, often stimulus-independent, release of dopamine, salience would be granted to what would otherwise be relatively innocuous events and stimuli. Without this salience events would not even become conscious and would not be attended to. Further, this aberrant salience is likely to mess up the temporal and contextual structure of experience.

It is argued that dopamine provides 'the wind of psychotic fire' (Laruelle & Abi-Dargham 1999). Salience enables objects to come to attention and demands an attitude from the organism and thus serves as a potent stimulus to goal-directed behaviour (Kapur et al. 2005a). In normal circumstances, context-driven activity of dopamine system mediates the experience of novelty and acquisition of appropriate motivational salience. In psychotic illness, there is a dysregulated dopamine system consequent upon genetic factors, early environmental insults, drug abuse, and 'social defeat' (Broome et al. 2005a). Such a system fires and releases dopamine independent of cue and context and thus events are accorded salience inappropriately (Kapur et al. 2005a). This neurochemical account offers parallels with an authorship account described above: the 'meaningful' connections derived are private: hence, when the deluded are asked to give reasons for the beliefs they either cannot, or if they do, we find it difficult to either view

Draft: Please do not quote without the authors' permission.

them as reasons or view them as poor reasons. For the subject with delusions, the dopaminergic-driven false attribution of salience generates the phenomenological state of private certainty with such pathological experiences being rationally interconnected and meeting normative demands. Yet, due to the private generation of meaning relations described above, these interconnections are not apparent to others who may interview the subject and such experiences, intersubjectively, appear as delusions.

Thus, psychosis may be viewed as a case of 'private language', albeit one that is always partially understood by one's interlocutors. Attention to the quality of reason giving and authorship is crucial when interviewing deluded subjects in psychiatry, especially in the early phase of psychosis. Clinical scenarios where such concepts are useful include those subjects who move from a prodromal or at-risk phase (Broome et al. 2005b; Yung et al, 2003) to a frank first-episode of illness, or when people with benign nonpathological unusual experiences may be referred for assessment. In these cases, eliciting the status of authorship and changes in reason giving may help provide a diagnosis. Those who retain authorship and justify their beliefs with reasons that others easily accept are more likely to be non-cases. The prodromal at-risk group, instead, are likely to show some lower status of authorship or offer reasons for their beliefs that do not maintain meaningful links with the belief contents according to the interviewer. These shifts may serve as an indicator of transition to psychosis and help determine when pharmacological treatment may be appropriate.

Draft: Please do not quote without the authors' permission.

The dopaminergic model of schizophrenia was, in the past, highly reliant on the fact that the central feature of the pharmacological treatment of schizophrenia lay in their effects on dopamine. Anti-psychotic medication works by blocking dopamine D2 receptors and hence, the continued inappropriate generation of salience and meaning. This effect may work relatively quickly (Kapur 2004; Kapur et al 2005b; Kapur et al. 2005a) but the unpicking of the pre-existing delusional system may require more time to resolve and may require psychological treatment, such as cognitive behavioural therapy. Thus, medication may stop the 'wind of psychotic fire' and hence the generation of new meaningful connections between representations, and further, minimize the affective hold or salience previously constructed relationships have over the subject. Medication, relatively quickly, will hence stop the elaboration of delusional systems and allow the subject to 'step back' and consider their beliefs more rationally. It is at this stage where psychological interventions may have a role to play (Garety et al. 2007; van der Gaag 2006). Here, the capacity for authorship will return and gradually the delusional contents will be incorporated into the subject's own account of their previous cognitive state: they may be forgotten, attributed to illness, 'stress', drug use or 'my mind playing tricks on me'.

Conclusion

The notion of authorship promises to help refine the current characterisation of delusions as pathologies of beliefs. First, it highlights the relation between rationality and self

Draft: Please do not quote without the authors' permission.

knowledge on the basis of the assessment of the capacity for reason giving. If the subject is unable to justify a belief, then she is not authoritative with respect to that belief and has not taken an agential perspective towards it. The content of many delusions is reported but is not justified on the basis of reasons and the delusion itself appears as badly integrated within the subject's belief system. Moreover, the subject often fails to act in a way that is compatible with endorsing the delusion. This indicates that no authorship is taking place and that the subject is not genuinely committed to the content of the delusions. To hypothesise a failure of self knowledge in some subjects affected by delusions has far reaching consequences: for instance, when subjects lose these capacities for belief authoring and belief integration, the intentionality of their delusional states and the degree of autonomy and responsibility they exhibit with respect to the thoughts that they report can also be doubted.

Second, attention to authorship is beneficial in diagnosis and in grading a subject's position along the continuum of psychosis, from healthy, to prodromal, to frank psychosis¹ and established schizophrenia. Engaging in the kind of questioning that allows one to establish authorship may also have ethical and clinical benefits: the subject may prefer to be interviewed this way as she feels her concerns are taken seriously; the therapeutic relationship is preserved as a relationship of mutual trust and genuine

¹ This refers to hallucinations or delusions of too short a duration to meet criteria for a diagnosis of schizophrenia as described in the ICD-10 and DSM-IV.

Draft: Please do not quote without the authors' permission.

exchange of information; and perhaps the subject can be persuaded to return to 'the space of reasons' in the course of some Socratic tutoring.

Finally, there is the explanatory challenge we started with. Can failure of reason giving be a promising demarcation criterion between delusional and non-delusional beliefs? Can authorship contribute to explaining the puzzling nature of delusions better than a more general appeal to irrationality? Reconstructing live debates in the philosophy of psychopathology around the notion of authorship promotes progress on a number of demarcation issues, including the nature of delusions and how they differ from framework beliefs. But this first exploration of the role of reason giving has revealed that some delusions are authored and some aren't, thereby lending support to the plausible view that there is no one kind of belief, no one kind of delusion, but a continuum of mental states that vary in content, and in potential for authorship, integration in the belief system and revision.

The emphasis on the quality and quantity of reason giving brings back complexity into the philosophical debate on delusions and highlights the interplay between different criteria for the classification of delusional states, on the basis of their content, their degree of elaboration, the certainty with which they are reported and especially the extent to which they are endorsed. The analysis of the puzzling nature of delusions does not rely on how many people find the content of the delusion plausible, but on the reasons (or lack thereof) that the subject uses it to justify it. Reason giving offers an external interpreter clues about the understandability, intentionality and rationality of the observed

Draft: Please do not quote without the authors' permission.

behaviour and offers the subject the opportunity to take control of the reported delusional state and manifest the endorsement of the delusional belief-content in a variety of situations. This is why we hope to see some empirical work investigating the quality and consequences of reason giving in psychopathology which would be valuable to the philosophical project of establishing a connection between rationality, self knowledge and autonomous decision-making.

References

- Andrenason, N. C. (2007). DSM and the Death of Phenomenology in America: An Example of Unintended Consequences. *Schizophr Bull*, 33: 108-112.
- Bayne, T. & Pacherie, E. (2005). In defence of the doxastic conception of delusion. *Mind & Language*, 20 (2): 163-188.
- Bayne, T. & Pacherie E. (2004). Bottom up or top down? *Philosophy, Psychiatry, & Psychology*, 11 (1): 1-11.
- Bermudez, J. (2001), Normativity and Rationality in Delusional Psychiatric Disorders

Draft: Please do not quote without the authors' permission.

Mind&Language, 16 (5): 493-457.

Berridge, K. C. & Robinson, T. E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Rev*, 28: 309-369.

Berrios, G.E. (1991). Delusions as 'wrong beliefs': a conceptual history. *British Journal of Psychiatry*, 159 (suppl 14): 6-13.

Bortolotti, L. & Broome, M.R. (2007). If you didn't care, you wouldn't notice: recognition and estrangement in psychopathology. *Philosophy Psychiatry, & Psychology*, 14(1): 39-42.

Bortolotti, L. (2005). Delusions and the background of rationality. *Mind & Language*, 20(2): 189-208.

Bortolotti, L. (2002). Marks of irrationality. In S. Clarke and T. Lyons (eds.) *Scientific Realism and Common Sense: Recent Themes in the Philosophy of Science* (pp. 157-174). Dordrecht: Kluwer.

Broome M. R., Johns L. C., Valli I., Woolley J.B., Tabraham, P., Valmaggia, L., Peters, E., Garety, P., & McGuire, P. (2007). Delusion formation and reasoning biases in those at clinical high risk for psychosis. *British Journal of Psychiatry* 191: s38-42

Broome, M. R., Woolley, J. B., Johns, L. C., Valmaggia, L., Tabraham, P., Gafoor, R., Bramon, E. & McGuire, P. (2005b). Outreach and support in South London (OASIS):

Draft: Please do not quote without the authors' permission.

implementation of a clinical service for prodromal psychosis and the at risk mental state.

European Psychiatry, 20: 372-378.

Broome, M. R., Woolley, J. B., Tabraham, P., Johns, L., Bramon, E., Murray, G.,
Pariante, C., McGuire, P. & Murray, R. (2005a). What causes the onset of psychosis?
Schizophr Res, 79: 23-34.

Broome, M. (2004). Rationality in psychosis and understanding the deluded. *Philosophy, Psychiatry, & Psychology*, 11(1): 35- 41.

Campbell, J. (2002). The Ownership of Thoughts. *Philosophy, Psychiatry and Psychology*, 9 (1): 35-39.

Campbell, J. (2001). Rationality, Meaning and the Analysis of Delusion. *Philosophy, Psychiatry, & Psychology*, 8 (2-3): 89-100.

Campbell, J. (1999). Schizophrenia, the space of reasons and thinking as a motor process. *The Monist*, 82(4): 609-625.

Carman, T. (2003). First persons: On Richard Moran's Authority and Estrangement. *Inquiry*, 46: 395-408.

Coltheart, M. (2005). Delusional belief. *Australian Journal of Psychology*, 57: 72-6.

Currie, G. (2000). Imagination, delusion and hallucinations. In M. Coltheart and M. Davies (eds.). *Pathologies of Belief* (pp. 167–182). Oxford: Blackwell.

Draft: Please do not quote without the authors' permission.

Davies, M., Coltheart, M., Langdon, R. & Breen, N. (2002). Monothematic Delusions: Towards a Two- Factor Account. *Philosophy, Psychiatry and Psychology*, 8(2/3): 133-158.

Eilan, N. (2000). On understanding schizophrenia. In D. Zahavi (Ed.) *Exploring the self* (pp. 97–113). Amsterdam: John Benjamins.

Ferrero, L. (2003). An elusive challenge to the authorship account. *Philosophical Psychology*, 16(4): 565-567.

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

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Science*, 4(1): 14-21.

Garety, P. A., Bebbington, P., Fowler, D., Freeman, D., & Kuipers, E. (2007). Implications for Neurobiological Research of Cognitive Models of Psychosis. *Psychological Medicine*, 37 (10): 1377-1391.

Gerrans, P. (2001). Authorship and ownership of thoughts. *Philosophy, Psychiatry and Psychology*, 8 (2-3): 231-237.

Graham, G. & Stephens, G. (1994). Mind and mine. In G. Graham and G. Stephens (eds). *Philosophical Psychology* (pp. 91-109). Cambridge (Mass.): MIT Press.

Draft: Please do not quote without the authors' permission.

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

Jeannerod, M. & Pacherie, E. (2004). Agency, Simulation and Self-identification. *Mind & Language*, 19(2): 113-146.

Kapur, S. (2004). How antipsychotics become anti-'psychotic'--from dopamine to salience to psychosis. *Trends in Pharmacological Sciences*, 25: 402-406.

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

Kapur, S., Arenovich, T., Agid, O., Zipursky, R., Lindborg, S., & Jones, B. (2005b). Evidence for onset of antipsychotic effects within the first 24 hours of treatment. *American Journal of Psychiatry*, 162: 939-946.

Kapur, S., Mizrahi, R. & Li, M. (2005a). From dopamine to salience to psychosis--linking biology, pharmacology and phenomenology of psychosis. *Schizophrenia Research*, 79, 59-68.

Laruelle, M. & Abi-Dargham, A. (1999) Dopamine as the wind of the psychotic fire: new evidence from brain imaging studies. *J Psychopharmacol*, 13: 358-371.

Lawlor, K. (2003). Elusive reasons: a problem for first-person authority. *Philosophical Psychology*, 16 (4): 549-564.

Draft: Please do not quote without the authors' permission.

Lewis, S. & Guthrie, E. (2002). *Master Medicine: Psychiatry*. Elsevier.

Moran, R. (2001). *Authority and Estrangement: an Essay on Self-knowledge*. Princeton: Princeton University Press.

Moran, R. (2004). Précis of Authority and Estrangement. *Philosophy and Phenomenological Research*, LXIX (2): 423-426.

Moyal-Sharrock, D. (2004). *Understanding Wittgenstein's 'On Certainty'*. Basingstoke, Hampshire: Palgrave Macmillan.

Mullins, S. & Spence, S. (2003). Re-examining thought insertion. *British Journal of Psychiatry*, 182, 293-298.

Parnas, J. & Handset, P. (2003). Phenomenology of Self-Experience in Early Schizophrenia. *Comprehensive Psychiatry*, 44 (2): 121-134.

Samuels R, Stich S & Bishop, M. (2002). Ending the Rationality Wars: How to Make Disputes About Human Rationality Disappear. In R. Elio (ed.), *Common Sense, Reasoning, and Rationality* (pp. 236-268). Oxford: Oxford University Press.

Sass, L. A. (1994). *The Paradoxes of Delusion*. Cornell: Cornell University Press.

Sims, A. (2003). *Symptoms in the Mind*. Saunders: Elsevier Publishing.

Draft: Please do not quote without the authors' permission.

Stanovich, K. E. (1999). *Who is Rational? Studies of Individual Differences in Reasoning*. Mahwah: Erlbaum Associates.

Stein, E. (1996). *Without Good Reasons*. New York: Oxford University Press.

Stephens, G. L. & Graham, G. (2006) The delusional stance. In M. Cheung Chung, W. Fulford and G. Graham (eds.) *Reconceiving Schizophrenia* (ch.10). Oxford: Oxford University Press.

Stephens, G. L. & Graham, G. (2004). Reconceiving Delusions. *International Review of Psychiatry*, 16(3): 236-241.

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

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

Thornton, T. (forthcoming). Why the idea of framework propositions cannot contribute to an understanding of delusion. *Phenomenology and the Cognitive Sciences*.

Tversky A. & Thaler R. (1990). Anomalies: Preference Reversals. *Journal of Economic Perspectives*, 4(2), 201-11.

Draft: Please do not quote without the authors' permission.

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

Velleman, J. (2005). The self as narrator. In J. Christman & J. Anderson (eds.), *Autonomy and the Challenges to Liberalism: New Essays* (pp. 56-76). New York: Cambridge University Press.

Wegner D. & Sparrow B. (2004). Authorship Processing. In Gazzaniga M. (ed.), *The Cognitive Neurosciences* (pp.1201-1209). Cambridge (MA): MIT Press.

Wilson, T. & Hodges, S. (1994). Effects of analyzing reasons on attitude change: the moderating role of attitude accessibility. *Social Cognition*, 11, 353-366.

Wittgenstein, L. (1969). *On Certainty*. Oxford: Blackwell.

Yager, J. & Gitlin, M.J. (2005). Clinical Manifestations of Psychiatric Disorders. In B. J. Sadock and V. A. Sadock (eds.) *Kaplan and Sadock's Comprehensive Textbook of Psychiatry Eighth Edition* (pp. 964-1002), vol. 1. Lippincott Williams and Wilkins, Philadelphia.

Yung, A. R., Phillips, L. J., Yuen, H. P., Francey, S.M., McFarlane, C.A., Hallgren, M. & McGorry, P.D. (2003). Psychosis prediction: 12-month follow up of a high-risk ('prodromal') group. *Schizophr Res*, 60, 21-32.

Draft: Please do not quote without the authors' permission.