Delusions and the Background of Rationality

LISA BORTOLOTTI

Abstract: I argue that some cases of delusions show the inadequacy of those theories of interpretation that rely on a necessary rationality constraint on belief ascription. In particular I challenge the view that irrational beliefs can be ascribed only against a general background of rationality. Subjects affected by delusions seem to be genuine believers and their behaviour can be successfully explained in intentional terms, but they do not meet those criteria that according to Davidson (1985a) need to be met for the background of rationality to be in place.

1. Introduction

In this paper, I am interested in elaborating and assessing an argument against those theories of belief ascription that rely on the rationality constraint (Davidson, 1974; Dennett, 1971; Heal, 1998). The rationality constraint on belief ascription (*RC*) is the view that one needs to be rational in order to be ascribed beliefs. RC has been often challenged in the literature on intentionality and belief ascription. Its critics have started from the observation that non-rational systems are often ascribed beliefs. Then they have attempted to turn these observations into counterexamples to those theories of belief ascription that explicitly rely on RC (see arguments in Stich, 1981; Lukes, 1982; Cherniak, 1986).

In order to reply to this kind of objections to RC, some philosophers (Davidson, 1982; Davidson, 1985a; Davidson, 1985b; Heal, 1998) have adopted the background argument (*BA*). According to this argument, only within a belief system that is largely rational can intentional description occasionally be granted to representational states that fail to meet the standards of rationality. Examples of beliefs that, according to this view, can be characterised in intentional terms only against a background of rationality are pairs of inconsistent beliefs, beliefs that result from reasoning mistakes or beliefs that are obviously false. The application of the background argument allows the interpreter to explain and predict the behaviour of non-perfectly rational systems in intentional terms.

How can we properly characterise the background argument? Does it make sense to talk about 'largely rational' belief systems? One influential characterisation,

I am grateful to Martin Davies, Kim Sterelny and Matteo Mameli for many helpful comments on earlier drafts of this paper. I would also like to thank two anonymous referees and the editor in charge of my submission for many useful suggestions.

Address for Correspondence: CSEP/IMLAB, School of Law, University of Manchester, M13 9PL, UK.
Email: lisa.bortolotti@manchester.ac.uk

Mind & Language, Vol. 20 No. 2 April 2005, pp. 189–208.

mainly due to Donald Davidson, is the following. A believer forms beliefs in a reliable way, is coherent, eliminates the tension among conflicting beliefs and revises beliefs in the light of new information. A believer can make reasoning mistakes and exhibit local inconsistencies, only because most of her beliefs have the right causal relations with the external world, are consistent with her other beliefs and cohere with the rest of her behaviour. There are two main requirements that a creature *A* must satisfy in order to count as a believer: (1) *A* cannot explicitly violate a fundamental norm of rationality and (2) *A* cannot fail to recover from a violation of a norm of rationality once this violation is made explicit to *A*. This capacity to restore rationality, together with the general conformity of the creature's behaviour to the standards of rationality, is regarded as sufficient for some version of RC to be in place.

In this paper I argue that the case of delusions can be constructed as a counter-example to BA as conceived by Davidson. My strategy will be to maintain that the belief system of deluded subjects can be plausibly described as lacking a general background of rationality. First, I shall introduce the notion of monothematic delusions (section 2). Then, I shall resist some arguments for the view that delusions are not beliefs, and emphasise that delusions can play the same role as beliefs in explanation, argumentation and action guidance (section 3). Once I have established that some delusions can be plausibly described as beliefs, I shall address the issue of whether belief systems affected by delusions lack a general background of rationality (section 4). I will attempt to show that the failure of rationality exhibited by deluded subjects is of the kind that prevents their behaviour from being characterised in intentional terms in the light of BA. Deluded subjects do not integrate their beliefs in a coherent system and might fail to revise beliefs when counterevidence against them becomes available. Moreover, deluded subjects might violate explicitly a fundamental norm of rationality (such as consistency) and often fail to restore conformity to the norm even after what Davidson calls 'Socratic tutoring'. In section 5, I shall assess one apparently promising reply that is available to the belief ascription theorist who endorses BA. The Davidsonian might concede that the beliefs of some deluded subjects can be described as non-rational, but insist that, thanks to BA, their behaviour can be characterised in intentional terms, because delusions are just another example of localised and temporary failures of rationality. I shall show that there are at least two reasons why this reply is not consistent with the best available reading of BA.

2. Delusions

I shall distinguish here between types of delusions. The term 'delusion' is used to refer both to the effects of brain damage (usually injuries that affect the right cerebral hemisphere) and to psychotic disorders with no known organic cause. In the former case, delusions are likely to be monothematic and circumscribed, while in the latter they are more likely to be florid and polythematic (Davies and Coltheart, 2000). These distinctions are relevant to my discussion.

A delusion is *monothematic* if the delusion is limited to one theme. For instance, a subject has a monothematic delusion if her reasoning seems otherwise unimpaired, but she systematically fails to recognise her image in the mirror and comes to think that there is a person identical to her following her around (mirrored-self misidentification). A delusion is *polythematic* if it extends to more than one theme, where the themes can be interrelated. For instance, a subject who believes she is surrounded by alien forces who control her own actions and are slowly taking over people's bodies might have an entire delusional system that is likely to affect the way she interprets most events occurring in her life. She both experiences 'alien control' and believes that the people around her are hostile (Payne, 1992).

Monothematic delusions might be more or less *circumscribed*. They are circumscribed if the content of the delusional states does not affect significantly the other beliefs and the behaviour of the subject. Monothematic delusions can be *elaborated*, if the subject draws consequences from her delusional states and forms other beliefs that revolve around the theme of the delusion. The same distinction applies in principle to polythematic delusions, though polythematic delusions tend to be elaborated rather than circumscribed.

In the rest of the paper, I shall focus on monothematic and relatively circumscribed delusions that are caused by brain damage. If deluded subjects explicitly violate norms of rationality and are impervious to recovery, then the adoption of RC prevents us from characterising their behaviour in intentional terms. On the other hand, if we do not describe their delusions as beliefs, we can hardly explain why subjects attempt to argue for the content of their delusion and why they often act on their delusion as they would act on their beliefs. The case of delusions offers us a *prima facie* reason to resist RC.

2.1. The Capgras Syndrome

Here I shall describe briefly the Capgras delusion, and refer to it in most of my later discussion. People subject to the Capgras syndrome claim that their spouses, or one of their close relatives, have been replaced by impostors. According to a widely accepted view (Stone and Young, 1997), the delusion arises when the affective component of the face-processing module is damaged, leaving recognition unimpaired. In the most typical case, the subject sees the spouse's face and recognises it, but forms the belief that the person is looking at is not really the spouse.[1]

On this interpretation, the delusion is an attempt to explain why the face seen, which appears identical to the familiar face, 'feels' strange (Gerrans, 2000). On most accounts, apart from an abnormal experience, a reasoning bias or a reasoning

[1] Further studies reveal that there is also an auditory form of the Capgras delusion, where the recognition of familiar voices is affected. This auditory impairment seems to be compatible with the hypothesis that the Capgras delusion is correlated with damages to the affective channel of recognition mechanisms. Reid *et al.* (1993) describe the puzzling case of a 32-year-old *blind* woman who experienced the Capgras delusion. She believed that her pet cat had been replaced by a replica that was hostile to her.

deficit is also necessary for the acceptance and the maintenance of the delusion (Langdon and Coltheart, 2000; Davies and Coltheart, 2000).

In most instances, the Capgras syndrome is a monothematic delusion, that is, it is often confined to the alleged substitution of the spouse or close relative with a clone, an alien or a robot that looks identical (or almost identical) to the original person. In some cases, Capgras patients act on their delusional beliefs by showing hostile or aggressive behaviour towards the alleged impostors. The character of delusions as action guiding seems to be occasionally supported by the tragic stories of some patients. One case often reported is that of Ms. A., who killed her mother after having suffered for five years the delusion that her parents were impostors (Silva *et al.*, 1994).

Pressed by questions, Capgras patients can also form other beliefs related in content to their delusional state in the attempt to explain how their delusion fits with other things they know. A patient was asked to explain why the 'impostor' had the ring he gave to his wife, and replied that the ring was not the same one, just a very similar one. Another subject was asked why he had not reported the disappearance of his wife to the police, and he candidly answered that the police would have never believed him. The more complex and articulated the explanations given by the subjects, the more elaborated is their delusional state, where elaboration is measured in terms of the number of connections that the delusional state has with other beliefs the subjects hold. Alternatively, subjects can go on with their lives substantially unchanged and show very little concern about their unusual situation and the alleged disappearance of their spouses or close relatives. In these latter cases, the delusion is circumscribed, that is, the delusional state does not necessarily interact with the patient's other beliefs, emotions and desires and does not necessarily lead the subject to act in accordance with it, probably as an effect of some radical form of compartmentalisation.

A general feature of both elaborated and circumscribed delusions seems to be the obstinacy with which subjects maintain the delusion in spite of its having a very implausible content and of other people's efforts to dissuade them. As I shall argue later, the extent to which resistance to change and compartmentalisation affect the subject's capacity to evaluate and reject her delusional state indicates that there is an element of irrationality in the subject's commitment to the content of her delusion.

3. Are Delusions Beliefs?

In this section, I shall focus on the issue of the nature of delusions. At least some delusions seem to play the same role that beliefs play. They are mental states grounded on experience and formed via an inferential process that involves beliefs accepted as true by the subjects. Delusions, as beliefs, are persistent and relate to other beliefs. Recall the case of the subject affected by the Capgras syndrome who was asked why he had not reported to the police his wife's disappearance, and answered that he didn't because he knew that police would not have believed him. In this case, the patient who believed that his wife had been replaced by an alien

also formed the belief that authorities would not listen to him because his story involved aliens. To some extent he had established connections between his delusional state and his other beliefs.

Delusions can lead to action as beliefs do. Both Freud (1917) and Young (2000) make sense of delusions as beliefs by stressing their action guiding character. Sometimes the actions performed by a subject affected by the Capgras delusion are the actions we would expect the subject to perform if she really believed the content of her delusion.

> It [the delusion] belongs first to that group of illnesses which do not directly affect the physical, but express themselves only by psychic signs, and it is distinguished secondly by the fact that 'fancies' have assumed control, that is, are *believed* and have acquired *influence on actions* (Freud, 1917, p. 173, my emphasis).
>
> Some of these [Capgras] patients do act in ways which can be seen as consistent with their delusion (Young, 2000, p. 49).

Most commentators agree with Freud and Young on the action guiding character of delusions, but some claim that resistance to counterevidence and circumscription speak against the identification of delusions with belief states. The search for a refined classification of delusions[2] is outside the scope of this paper. My purpose is to argue that some delusions *can be plausibly described as beliefs*, and so I shall attempt to resist some influential arguments for the view that delusions are not beliefs. I shall conclude that some of the arguments against the claim that delusions are beliefs rely on a narrow conception of belief.

The main contention is that the role of a mental state in *reality testing* is what determines whether a mental state is a belief state. The acceptance of the belief must be responsive to evidence, to how reality is. The subject must be prepared to give up the belief if evidence against it becomes available. This does not seem to happen in the case of delusions. Berrios (1991) maintains that when the subject is deluded, she is not expressing a belief, but is using 'belief-talk' to make sense of the strange experience she has. Sass (1994) also argues that delusions do not meet the criteria for beliefs, because they are not sensitive to how reality is. The subject would not even mean to refer to the real world with her delusional state, but just to the private world of her experience. The world of the subject's experience would not be, and would not be thought to be, even by the subject, the real world.[3]

---

[2] Here is one of the *official* definitions of delusions that grant them the status of beliefs. 'Delusion. A false belief based on incorrect inference about external reality that is firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary'. From the Diagnostic and Statistical Manual of Mental Disorders, American Psychiatric Association (1994), p. 765.

[3] The view that the deluded subject might be just using belief-talk to express her delusional state seems to make sense in those cases in which the delusion is heavily compartmentalised and does not have any significant impact on the subject's behaviour.

Berrios' position is widely discussed in the literature on the nature of delusions. Apparent evidence for his account is that the subject often recognises the absurdity of the content of her delusion, but that is not a sufficient reason for her to abandon the delusional state (Young, 2000). This is how Berrios argues against the view that delusions are beliefs:

> We must now test the hypothesis that, from the structural point of view, delusions are, in fact, beliefs. Price (1934) distinguished four elements that comprise a belief (P):
>
> a) Entertaining P, together with one or more alternative propositions Q and R;
> b) Knowing a fact or set of facts (F), which is relevant to P, Q or R;
> c) Knowing that F makes P more likely than Q or R;
> d) Assenting to P; which in turn includes (i) the preferring of P to Q and R; (ii) the feeling a certain degree of confidence with regard to P.
>
> Price's criteria are clear and elegant enough, but it is clear that no current textbook or empirical definition of delusion can be set in terms of these four criteria (Berrios, 1991, p. 8).

Berrios argues that delusions do not meet the criteria (a)–(d) for beliefs and that we should not regard them as beliefs. Berrios seems to offer two main reasons for that. First, rarely would there be a fact or a set of facts that is relevant to the delusion and supports it (see point b). Secondly, even if there were a fact or set of facts in need to explanation, the subjects would not assign higher probability to the content of the delusion than to alternative hypotheses (see points c and d). Independently of the plausibility of Price's criteria, Berrios' worries can be reformulated in more general terms as genuine arguments against the view that delusions can be plausibly described as beliefs, and so they deserve our attention.

3.1. The Formation of the Delusion

Let's consider the problem of the formation of the delusion first. One way of accounting for what happens to subjects affected by Capgras is the following. Subjects adopt the view that their spouses or close relatives have been replaced, because there is a *fact* that is relevant to the content of their delusions, i.e. the fact that, all of a sudden, their spouses or close relatives look different to them. If there is a fact in need of explanation, subjects have a *reason* to claim that their spouses or close relatives have been replaced. It is true that their reason to form the delusional belief might not be a *good* reason, and that the belief should be abandoned on the basis of powerful considerations against it (e.g. implausibility). However, this is relevant to the assessment of the plausibility of the delusional state and not to the issue whether the delusional state is a belief. It is a triviality that not all our ordinary beliefs are justified, held for good reasons or plausible.

Some object to this explanation of what happens to deluded subjects.[4] The account I have sketched seems to require that the subjects' failure to feel an emotional response towards their spouse or close relative be conscious and accessible to them as a reason for their beliefs (*personal* account). It is important to distinguish the view that subjects are aware that they lost their affective response to the familiar face from the view that subjects are aware that the familiar face looks different. The personal account I favour claims that it is possible that the deluded subjects consciously experience that the spouse or relative looks different and come to believe that this is the case. Deluded subjects are not necessarily aware of the fact that the affective response is missing. We do not seem to be aware of our affective responses in the great majority of cases, and sometimes only skin conductance can reveal whether the response is present. But it is sufficient that the subjects are aware that the face looks different to them, that something is wrong, in order for them to have *something* to explain.

What would be an alternative account of the formation of the delusional state? The content of the subjects' experience would be neither consciously accessed by the subject nor analysable. The delusional state would be formed as an effect of the subjects' experience, but the experience would not lend any personal-level support to the content of the delusion. The delusion would be the 'output of a modularized affective subsystem' (Gerrans, 2000, p. 115) and therefore the subject would have no conscious access to the reasons for the formation of the delusion. I shall call this alternative explanation the *subpersonal* account. Evidence for it is that the subject cannot avoid forming the delusion, and has trouble revising it, even if she recognises how implausible it is.

There are two issues related to the consideration of the subpersonal account of the formation of the delusion. One is whether the account is defensible, and the other is whether it is compatible with the identification of delusions with belief states. It seems to me that the account can be defended, if empirical data in its support emerge, though we have no reason at this stage to think that it is more plausible than its alternative. However, even if the subpersonal account did prevail in light of future research, it would not necessarily speak against the view that delusions are beliefs.

At first glance, the subpersonal account does not square with the data we have about the subjects' attempts to justify their behaviour. When they insist that their spouse or relative looks different, they are also trying to persuade their interlocutors that they are right. The presumed difference in looks or behaviour between the real spouse or relative and the 'impostor' becomes *evidence* for the delusion. Often subjects attempt to describe the difference between the impostor and the spouse or relative in terms of slight discrepancies in the physical appearance, such as 'the eyes are too close' or 'she's too tall' (Breen *et al.*, 2000).

---

[4] Tony Stone in conversation and in 'Implicit and Explicit Processes in Delusional Belief Formation', presented at the AAP conference in Christchurch, New Zealand, 2002.

The obvious response to this line of argument is that the attempt to justify one's conviction is just a *post-hoc* reconstruction. The physical difference might be something to which the subject appeals in order to rationalise her commitment to the implausible content of the delusion. According to this view, once the subjects find themselves with the delusion, they try to make it sound plausible to themselves and others by reference to perceptible differences between the physical appearance of their spouses or relatives and that of the alleged impostors.

Another reason to reject the idea that the content of the delusion is not analysable, is that the content of the delusional state seems to track the popular culture of the times. People affected by the Capgras delusion in the 70s, when the existence of UFOs was much discussed, tended to claim that extraterrestrial beings were responsible for replacing their spouses or relatives. Nowadays, subjects tend to identify the 'impostor' with a clone of the original. The presence of these details in the explanation of the phenomenon of substitution and their changing over time, seem to indicate that the delusional state is a belief consciously formed in order to make sense of a puzzling experience. But, in the same fashion as before, the defender of the subpersonal account might just argue that these considerations come *after* the formation of the delusional state and are supposed to play at most an explanatory and justificatory role.

If we accept the personal account, then it is hard to argue that delusions are not beliefs. The delusion would be formed in the same way as paradigmatic beliefs are, according to Berrios, as an explanation for a fact that is relevant to the subject. If we opt for the subpersonal account, it is still not ruled out that delusions can be described as beliefs. There seems to be nothing intrinsically incoherent in the view that some beliefs are not the product of personal-level hypothesis formation. The endorsement of the subpersonal account is not by itself sufficient to deny that delusions are beliefs. Arguably, our everyday perceptual experience is also the effect of a modularised system and lends support to our beliefs in a retrospective fashion. A typical perceptual belief is not produced through a conscious process of hypothesis formation. It isn't the case that one sees a chair and, by a conscious process, one forms the hypothesis that there is a chair. Rather, what happens is that the perception of the chair gives rise to the perceptual belief that there is a chair independently of any conscious hypothesis formation. Once the perceptual belief has been formed, one can justify and rationalise one's belief that there is a chair by appealing to one's perceptual experience of the chair. In this way, one reconstructs the belief formation process in terms of the conscious adoption of hypotheses that are explanatory of one's experience.

3.2. Delusions as Imaginings

Let's turn to the issue of probability assignments. Berrios seems to think that, when we believe that *P*, part of what it is to hold that belief is to regard *P* as more probable than some relevant alternatives *Q* and *R*. If the content of the delusion is not regarded as probable by the subject, it cannot be said to be *believed* by the subject and could just be the content of an act of imagination. It is true that the subject who believes his wife has been replaced by aliens and doesn't go to report

the fact to the police, *knows* that the content of his delusional state is not probable. He knows that it is not something that other people would easily believe. Still, we can all imagine situations in which people end up believing things that they do not consider more probable than their relevant alternatives. An example is that of religious beliefs. Any Christian would agree that a man's resurrection is less probable than his dead body being stolen. Still, Christians believe that Jesus rose from the dead and that his body was not stolen.

More to the point, consider this fictional case. Last night Liam could not sleep and decided to get some fresh air. He went out in the back garden, was struck by a sudden light and when he could open his eyes again, he saw a little green man in front of him. Today he tells his friend the story and says: 'I know that meeting little green men in one's garden is not a common thing and, if you had told me this story, I would have not believed you. But I *did* see a green man'. The fact that Liam saw a the little green man is a prima facie reason for him to believe that there was a little green man. Further considerations about the possibility that he might suffer from hallucinations or that he might have been the victim of a joke could then contribute to his giving up the belief that there was a little green man in his back garden. Those alternatives could become relevant at a later stage. Deluded subjects might be in a similar position to that of Liam in the story, that is, they might have evidence for a very improbable belief and neglect considerations about relevant alternatives. This is not unlikely, given that the relevant alternative to believing something improbable, for subjects affected by delusions like the Capgras syndrome, is to accept that there is something very wrong with them.

There is another influential account of delusions in the literature that deserves to be mentioned, even though it has not been developed to explain delusional states such as the Capgras syndrome. Currie (2000) suggests that the delusion of a subject *A* is due to *A*'s misidentifying *A*'s own attitude to some propositional content. For instance, let's suppose that Jim suffers from persecutory delusions. He imagines that everybody hates him. What makes him deluded is that he misidentifies his imagining as a belief. So Jim takes himself to *believe* that everybody hates him and acts in accordance with the belief he takes himself to have. In this particular context, Jim is not a good interpreter of himself and, as Currie puts it, there is something amiss in his meta-representational capacities.

Is Currie's suggestion compatible with my attempt to characterise delusions as beliefs? According to Currie, there is a belief involved in the delusion (Jim's meta-belief that he believes that everybody hates him) and that belief does not seem to be rationally held. If I rationally believe that I believe that *p*, it is because I have reasons to believe that I believe that *p*. But Jim does believe that he believes that *p* without in fact having any reasons for believing that he believes that *p*.

Jim's behaviour can be explained and predicted via the ascription of a non-rational belief, the meta-belief that he believes that everybody hates him, whereas he just imagines that this is the case. What is responsible for the abnormality of his delusional state is not what Jim imagines. It might be perfectly acceptable for Jim to imagine that everybody hates him. What is wrong with him is the fact that he takes himself to believe what in fact he just imagines. On this account, just like on my

account, there is a non-rational belief involved in the delusion. For Currie, this is a special kind of belief, a belief about what one believes. Moreover, Currie concedes that, when a subject believes that she believes that *p*, often she also ends up believing that *p*. So, going back to my fictional example, Jim might end up believing that everybody hates him as an effect of his believing that he has that belief. On the basis of these considerations, even those delusions that might be plausibly characterised as imaginings misidentified as beliefs can be seen as involving non-rational beliefs.

One could argue for the thesis that delusions are not beliefs in different and apparently legitimate ways, because there is no consensus on the necessary and sufficient conditions for what beliefs are. 'Belief' is a term used to refer to mental states with variable source and persistence, and incredible diversity of content. Central features of beliefs, such as their revisability or their action guiding character, under close examination fail to be necessary conditions, as there are so many examples at hand of beliefs that do not exhibit those features. That is why there cannot be any conclusive arguments to the effect that delusions are beliefs. Here, I have just attempted to resist some reasons for denying that they are.

According to the best available account of the formation of delusional states, the delusion is an attempt to explain a puzzling experience the subject has or had. Once formed, the delusion is often defended by the subject with tentative arguments. In this respect, delusions do not differ from ordinary beliefs. I have also observed that, although the delusion is not the most probable explanation for the subject's experience, delusions are not the only beliefs that are maintained in spite of not offering the most probable explanation of the relevant facts. In particular, in the case of delusions, the relevant alternative is a hypothesis that the subject has a strong motivation to resist.

For the rationality constraint theorist, we ascribe beliefs to others on the basis of interpretation. The beliefs we ascribe to the subject must rationalise the subject's behaviour and help us predict her future actions. Some have argued that delusions do not play the same role that paradigmatic instances of beliefs play in reality testing, as the acceptance of the delusion is not responsive to evidence. However, delusions have many of the features that paradigmatic instances of beliefs have, such as their action guiding character and their capacity to relate to other beliefs in inferences. From the point of view of an interpreter, the ascription of a delusional state to a subject might turn out to be the only way to make sense of the subject's behaviour (for instance, her sudden hostility towards a loved one). The ascription of the delusion, just like the ascription of beliefs, contributes to a better understanding of the intentional systems around us.

4. The Irrationality of Delusions

In order to have a counterexample to BA, I need to show that there are belief systems that lack a general background of rationality on the basis of the criteria described in

section 1. Deluded subjects are not irrational in every respect, but might fail to integrate beliefs in a coherent system and to revise beliefs when counterevidence against them becomes available. To challenge BA, I need to argue that deluded subjects can explicitly violate fundamental norms of rationality and fail to restore rationality even when their attention is drawn to the violation of the norm. In the section above, I have suggested that there are no good reasons to deny that delusions are beliefs. Here, I have to show that systems affected by delusions have those characteristics that prevent them from being described as largely rational in the light of BA.

While delusions seem to be a paradigmatic case of irrationality, theorists disagree about the sense in which delusions can be said to be irrational. My suggestion is that there is no principled way to distinguish delusions from other non-rational beliefs (Bortolotti, 2002). A belief (delusional or otherwise) might fail to be rational for one or more of the following reasons:

1. The belief is formed without there being sufficient or adequate justification for it;
2. The belief is compartmentalised, that is, it does not cohere with other beliefs that belong to the same system and with the rest of the subject's behaviour;
3. The belief is maintained in the face of strong counterevidence.

These three possible failures of rationality roughly correspond to three processes that characterise any belief system: formation, integration and revision of beliefs. Most authors identify the irrationality of delusions with failures at the level of coherence and revisability (points 2 and 3).

> Rationality is a normative constraint of consistency and coherence on the formation of a set of beliefs and thus is prima facie violated in two ways by the delusional subject. First she accepts a belief that is incoherent with the rest of her beliefs, and secondly she refuses to modify that belief in the face of fairly conclusive counterevidence and a set of background beliefs that contradict the delusional belief (Gerrans, 2000, p. 114).

> Delusions do not seem to respect the idea that the belief system forms a coherent whole and that adjustments to one belief will require adjustments to many others (Young, 2000, p. 49).

Ideally, we would want our belief systems to include mostly well-justified beliefs and to exclude beliefs that are in tension with the available evidence or with other justified beliefs in the system. The beliefs in the system ought to be well-grounded, responsive to evidence and coherent with one another. This is by no means supposed to be an exhaustive picture of the rationality of beliefs, but rather a rough guide to the detection and classification of deviations from rationality. Now I shall briefly explain how delusions fare in relation to the three criteria mentioned above.

4.1. Belief formation, Integration and Revision

Maher (1974b) has argued that delusions are not ill-formed beliefs. The abnormality of the delusion would be entirely due to the abnormality of the experiences on the basis of which the delusion is formed. Maher argues that that deluded subjects do not violate normative standards of belief formation more often than other believers do. The only significant difference between ordinary believers and deluded believers is that the latter have to account for abnormal experiences caused by brain damage. In the case of the Capgras, the abnormal experience is the perception that the face of the spouse or relative is different, and it is due to the fact that the subject does not have any affective response to the face. The content of the abnormal experiences is responsible for the implausible content of the delusional states.

Once we accept that some version of the personal account of the formation of the delusion is true, there is still the problem of understanding why deluded subjects 'choose' to account for their experience by accepting such implausible beliefs. Several explanations have been attempted. Some psychologists suggest that it is the subjects' state of paranoia and suspiciousness that makes them attribute a change in the external world rather than in themselves. This would be due to an attributional bias generated by motivational factors. An alternative view is to claim that the subject's reasoning is affected not by a motivational bias, but by a deficit at the level of belief evaluation. On this view, the deficit is permanent and is manifested in all cases of belief formation and not just those that may have a special motivational salience. Empirical investigation will contribute to the debate on whether deluded subjects have a motivational bias or a reasoning deficit, but it is not clear at this stage how the account according to which a permanent deficit is responsible for the delusion would explain the occasional recovery of some subjects.

Maher's suggestion that belief formation is not affected in deluded subjects is an interesting one, but cannot have the consequence of establishing that delusions are not irrational. Subjects might form their delusional belief in perfectly legitimate ways, but why do they *accept* the belief and *maintain* it? One way of making this point is to refer to the analogy with the case of everyday perceptual illusions. We see the straw in the glass of water as bent or broken. We might even come to believe—for instance the first time that we are subject to the illusion—that the straw *is* bent. But then we revise our belief when we realise that our visual experience was misleading. Capgras patients have many good reasons to doubt the content of their experience and of their delusional belief, including the authority of their therapists and the testimony of people they trust, but they do not easily give up their delusional belief once they have formed it.

The resistance to change in the delusional belief is particularly surprising because the subject's sensation that something in the spouse or relative has changed often evaporates in situations in which the subject does not see the face of the spouse or relative. For instance, when subjects hear the spouse or relative's voice on the phone, they have no trouble recognising it. This phenomenon should have an

effect on the Capgras patient that is similar to the effect of viewing the straw out of the glass after having been subject to the perceptual illusion that the straw was bent or broken. In the case of illusions, after seeing that the straw is not bent, we realise that experience had misled us to believe that the straw was bent and we give up the belief that it is. But the Capgras patients, who hear the same voice both when they talk on the phone with their spouse or relative and when they talk to them in person, recognise them in the former case and do not recognise them in the latter.

In general, we revise a belief that we have formed as a consequence of having a certain experience, if we have reasons to doubt the content our experience or if the belief strikes us as implausible given other things we know, or given some other experiences we have. This is how rational belief systems operate. The suggestion here is that rationality does not concern belief formation alone (Langdon and Coltheart, 2000, p. 190; Leeser and O'Donohue, 1999, p. 690). The appeal to a strange experience can explain why a false belief is formed, but not why it is accepted and maintained irrespectively of counterevidence and lack of support from other beliefs.

That is where some emphasis on the process of belief integration can be helpful in identifying the reasons why delusional states deviate from rationality. As I have anticipated, the appearance of the delusion generates inconsistencies that the Capgras patients are not ready to give up, despite people around them draw their attention to the inconsistencies. Breen interviewed a patient affected by a delusion of misidentification, RZ, a 40-year-old woman. RZ was asked who she was. First, she answered this question by giving her father's name, and then her grandfather's. In the course of the same interview, asked what her sex was, she claimed to be a man and then a woman, in spite of the interviewer's attempt to make her realise that she was being inconsistent. It is important to note that, during the interview, RZ was perfectly capable of defending her claims. She was ready to follow each of her answers with some explanation of why she was what she claimed she was and to give some evidence for her claims. For instance, when she claimed she was a man, she added that her facial hair was evidence for that—the patient had abnormal facial hair as a result of hormonal dysfunction (Breen *et al.*, 2000, pp. 93–98). Other interesting cases of inconsistent behaviour in deluded subjects are presented by Stone and Young (1997).

After reading these reports, one is tempted to concede that there might be believers who fail to recover from inconsistencies, even when they are questioned about the subject of their inconsistency. There seem to be circumstances in which the deluded subject does not give up a pair of inconsistent beliefs in spite of being subject to 'Socratic tutoring'. If this is right, then, some deluded believers violate one of the conditions imposed by BA, that is, they do not recover from violations of fundamental norms of rationality.

The main reason why delusions are perceived as non-rational beliefs is due to their being resistant to revision in the face of powerful counterevidence. It is at least to some extent true that, as Berrios says, delusions are excluded from the game of reality testing. They resist revision in an anomalous way and have also been

defined as beliefs held in the face of evidence normally sufficient to reject them. In a typical scenario, the Capgras patient comes to believe that, say, her mother has been replaced by an impostor. The subject holds on to her belief, even though the 'duplicate' acts exactly like her mother, has her mother's memories and is recognised as her mother by other relatives and friends. Moreover, a figure of authority, typically a psychiatrist, informs the subject that she is suffering from a rare but well-known syndrome due to brain damage. Yet, the subject's belief that her mother has been replaced by an impostor is unshaken.

However, resistance to change is not an exclusive character of delusions and is not always a mark of irrationality.[5] My point is that the perseverance that is typical of delusional states is a characteristic of some ordinary beliefs as well. Think about the case of a hinge proposition, that of a scientific theory strenuously defended in spite of counterevidence, or everyday cases of self-deception. Delusions are anomalous, but their anomaly is due to the *extent* to which they are resistant to change, and not to the very fact of their resistance. Those who claim that only delusions are subtracted to the game of reality testing ignore some interesting phenomena that concern ordinary beliefs.

To sum up, delusions can be formed for reasons that are not good reasons and are accepted even though there might be overriding considerations against their acceptance. Moreover, some systems affected by delusions end up being radically compartmentalised, resist change in presence of strong counterevidence and might fail to recover from violations of fundamental norms of rationality, such as inconsistencies. These cases suggest that some deluded subjects do not meet the conditions set by BA for being ascribed beliefs.

5. Assessing the Damage to the Rationality Constraint

To what extent does the case of delusions undermine RC? I believe the case of delusions is problematic for anyone who wants to defend the rationality constraint by means of BA. According to BA, it is possible to ascribe beliefs to systems that are not perfectly rational. In any belief system a background of rationality can be found against which local deviations can be explained. Davidson adds two conditions that systems must satisfy in order to be regarded as largely rational. They cannot violate explicitly a fundamental norm of rationality and they must be able to recover from violations of norms of rationality once these are made explicit.

I have argued that monothematic delusions can be seen as intentional states that make sense of subjects' behaviour but do not meet the conditions for intentional characterisation as set by BA. Subjects affected by monothematic delusions fail to

[5] Resistance to powerful counterevidence might be. By 'resistance to powerful counterevidence' I mean the phenomenon according to which a belief content remains unaltered, even though the reasons why the belief was accepted cease to provide adequate justification for it and new evidence against that belief becomes available.

recover from violations of fundamental norms of rationality even when they are made aware of such violations.

5.1. A Counter-objection

Given what we know about delusions, a theory of belief ascription should be able to account for the intentional characterisation of delusional states. Now the question is, can a Davidsonian account for delusions as genuine beliefs? I shall consider here a reply that focuses on the case of monothematic and relatively circumscribed delusions.

The resourceful Davidsonian has one prima facie plausible reply to the argument from delusions: a good case to be made for the application of BA to some deluded subjects. Instead of arguing that delusions are not beliefs, one could argue that belief systems that include delusional states can preserve a general background of rationality. We can describe delusional states in intentional terms just because the behaviour of deluded subjects *is* largely rational and it can be interpreted via the ascription of beliefs that are largely true and consistent. If delusions like the Capgras could be seen as temporary and localised irrationalities in a system that is largely rational, they should not be treated differently from occasional reasoning mistakes or evidently false beliefs. In the case of relatively circumscribed and monothematic delusions, the belief system in its totality is not 'contaminated' by the irrationality of the delusional belief and the deluded subject can be perfectly rational when the topic of the delusion is not raised.

There are two facts that seem to lend some support to this response. First, as we have already seen, subjects often appreciate the implausibility of the content of their delusion and therefore seem to maintain a critical attitude towards it. Secondly, some subjects can recover from their delusional beliefs and therefore their failure of rationality could be seen as temporary. These two phenomena, the maintenance of a critical attitude towards the content of the delusion and the possibility of recovery, might speak in favour of the application of BA to subjects affected by monothematic and relatively circumscribed delusions.

However, there are some considerations that speak against the application of BA to these cases. Whether deluded subjects have belief systems that are largely rational depends trivially on what we mean by 'largely rational'. On the best reading of this phrase, I shall argue, belief systems affected by monothematic delusions do not qualify as largely rational.

In the first section, following Davidson, we have defined a subject's behaviour as largely rational when *most* of the subject's beliefs have the right causal and evidential relations with the world and are in the right causal and logical relations to each other. The question is whether the belief system affected by delusions is largely rational in this sense. It is important to establish what factors are relevant to the assessment of rationality according to this approach. Is it relevant *how often* the deviations from rationality occur? Is *the extent* to which the behaviour deviates from rationality also an important factor? Consider the following example.

Whenever Martha needs to test a conditional statement, she checks whether the antecedent is true. She commits the same mistake that most experimental subjects make when asked to solve the selection task. Bob believes that the person who lives in his house and looks identical to his mum is not his mum, but a cleverly disguised Martian. Suppose Bob and Martha are otherwise perfectly rational. Is there a sense in which Martha is more rational than Bob?

Let's see whether we can distinguish between these cases. For one thing, Martha's failure of rationality is a very common one, whereas Bob exhibits a very rare condition, but this seems not to be a relevant factor in our present discussion. In my view, we can detect a fairly clear distinction when we examine in more detail the behaviour of the two. Bob's belief integration and revision processes seem not to be operating as well as Martha's.

Bob's delusion is likely to resist integration and revision more strenuously than Martha's ill-formed belief. Suppose Martha's logic tutor explains to her that, in order to test the conditional statement, she needs to check whether the consequent is true, and he offers her some examples of application of the rule. Then, she will come to believe that she had made a mistake. She will be disposed to recover from her failure of rationality when this is explained and pointed out to her. When the psychiatrist tells Bob that there are no Martians and that, even if there were any, they probably would not kidnap people and replace them with seemingly identical substitutes, he might recognise that what happened to his mother is incredible. However, it is very unlikely that he will give up his delusional belief, in spite of the evidence the psychiatrist and others might gather against it and the alternative explanations they can offer for the fact that his mother looks different to him.

The only reading of BA that would make it true that monothematic and circumscribed delusions can be intentionally characterised is the reading according to which a belief system is largely rational if most of the beliefs it contains are rational. This reading lacks the resources to discriminate between the behaviour of believers who can recover from failures of rationality and those who cannot, a distinction that plays a very important role in Davidson's characterisation of BA. The simplistic reading cannot therefore be the best available reading of BA. It is not consistent for the Davidsonian to illustrate the role of BA by appealing to the phenomenon of recovery *and* to endorse the view that deluded subjects are largely rational. For Davidson, it is a condition on believers that, after 'Socratic tutoring', they recover from deviations from rationality. Deluded subjects typically don't.

There is a further reason why I believe it is not in the spirit of BA to consider the behaviour of people affected by relatively circumscribed and monothematic delusions as largely rational. The idea behind BA is that the non-rational belief can be made sense of only *by appealing to the other true and sensible beliefs* the subject has. That a false belief can be explained by appealing to the other true and justified beliefs that belong to the same system, is often true. Heal endorses some version of BA and explains how mistakes can be rationalised:

> [. . .] when a mistake is agreed to have been made we will often look for, and find, a reason why it was made, not just in the sense of cause or regularity in its making but in the sense of some excuse which reconciles the mistake with the idea that, even in making it, the perpetrator was exercising his or her rationality (Heal, 1988, p. 99).

It is not a coincidence that BA has been devised to deal with certain types of failures of rationality. It can offer a good framework to understand how people can commit explicable errors and end up endorsing a false belief as a consequence of not grasping or mastering a concept.

Suppose that Jennie's belief that there is an elm in front of her is false: the tree in front of her is a beech. We can make sense of Jennie's mistake though, because it is an explicable one. She saw a tree and thought it was an elm. She rightly believes that elms are trees, and that there is a tree in front of her, but she does not realise that the tree in front of her is actually a beech.

Think about the case of Bert who goes to the doctor and claims he has arthritis in his thigh. We can understand what happens here: Bert has a pain in his thigh and does not realise that arthritis is a condition of the joints only. He does not completely master the concept 'arthritis' (or his concept 'arthritis' is not the same as his doctor's concept 'arthritis').

Finally, imagine Bill is helping his dad to fix the car. Bill's dad raises his head and asks Bill to pass him the sandals. Bill notices that there is a pair of sandals in the garage, but also realises that what his dad wants is the spanner. In fact, when Bill does pass him the spanner, his dad thanks him and goes on fixing the car. Bill's dad had a slip of the tongue.

There are strong analogies between these cases. The speaker makes a mistake that the interpreter can understand. It is open to the interpreter to apply the principle of charity and make sense of the fact that Jennie is referring to a beech, Bert is talking about a pain in the thigh and Bill's dad wanted the spanner. This is possible because we know that Jennie and Bert would not be resistant to revise their statements if an 'expert', or an otherwise suitably located interpreter, were to explain to them, respectively, the difference between elms and beeches and the exact meaning of arthritis. We also realise that Bill's dad would recognise his mistake if it were pointed out to him. These cases are the ones that BA is best suited to account for. The interpreter makes sense of what is going on by reference to the evidence that the speaker has for what she believes and to her behaviour in the context.

The same strategy cannot be easily applied to delusional states. Which true belief of Bob will help us rationalise his conviction that the person who looks identical to his mother is actually an alien? In what sense is Bob's mistake explicable? If Bob's belief can be rationalised, it is by reference to the strange experience to which he is subject, but that, as I argued, does not go far in explaining why Bob hangs onto the belief.

Moreover, trying to reinterpret Bob's belief in order to restore rationality seems hopeless. As a charitable interpreter, I could take Jennie as saying: 'There is a beech

in front of me', instead of 'There is an elm in front of me'. How could I reinterpret Bob's belief? If I interpreted Bob as having the belief that he is suffering from the Capgras delusion, I could not make any sense of his behaviour, especially of his hostile behaviour towards his mum. My charitable ascription would not be helpful in attempting to explain and predict Bob's behaviour in intentional terms.

If the idea behind BA is plausible at all, it is plausible in those cases in which the believer makes a revocable mistake or is in partial ignorance. In other cases, the idea that we are better off when we take the believer to endorse something true and sensible seems not to be supported by our experience as everyday interpreters of others. We seem to ascribe beliefs to people who make reasoning mistakes and are inconsistent or delusional, just because we allow the same possibility that RC denies, that the person in front of us might not be rational and might not tell us something true. It is the role that a presumed belief-state plays in the system that leaves the interpreter no choice as to treat that state as a candidate for intentional description and that system as a belief system. Beliefs typically relate to other beliefs, respond to the system's concerns, are manifested in behaviour and guide actions.

6. Conclusion

In this paper I have challenged the background argument that has been developed in order to defend the rationality constraint on belief ascription from potential counterexamples. The background argument rules out that some subjects can be ascribed beliefs, in particular those subjects who violate fundamental norms of rationality and fail to restore rationality when explicitly invited to do so. Deluded subjects do not satisfy the requirements of the background argument, but we seem to ascribe to them beliefs that make sense of their behaviour.

I have argued that we are better off if we ascribe beliefs to at least some of the people who suffer from delusions, because their behaviour can often be successfully explained and predicted within the intentional stance. They behave as believers when they act on their delusion, defend the content of their delusion with simple arguments and relate the content of their delusion to other beliefs they hold. Their behaviour is 'anomalous' in that they resist counterevidence and their delusional states are partially compartmentalised in their belief systems. But resistance to change and compartmentalisation are features often shared by non-delusional beliefs.

A rationality constraint theorist impressed by my defence of delusions as beliefs might be tempted to regard delusional states as an example of those deviations from rationality that can be accounted for on the basis of the background argument: temporary and localised. In this case, she could concede that delusions are non-rational beliefs and explain that it is possible to characterise them intentionally because they belong to a system of beliefs which is otherwise rational. This option

is not open to the rationality theorist given the best reading of the current formulation of the background argument. What is special about delusions is the extent to which they might be resistant to change or compartmentalised and the fact that they can exhibit both features at once. Deluded subjects might deviate from fundamental norms of rationality and fail to restore rationality when pressed to do so. These aspects of their behaviour suggest that they do not meet the conditions for intentional characterisation set by the background argument.

The case of delusions provides a reason to challenge the general idea that there must be a necessary connection between the intentionality exhibited by believers and rationality. In particular, the background argument seems to be unsuited to deal with the possibility of the intentional characterisation of delusional states.

Centre for Social Ethics and Policy
School of Law
University of Manchester

References

Bentall, R.P. *et al.* 2001: Persecutory delusions: a review and theoretical integration. *Clinical Psychological Review*, 21 (8), 1143–1192.

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

Blackwood, N.J. *et al.* 2001: Cognitive neuropsychiatric models of persecutory delusions. *American Journal of Psychiatry*, 158 (4), 527–539.

Bortolotti, L. 2002: Marks of irrationality. In S. Clarke and T. Lyons (eds.), *Recent Themes in the Philosophy of Science*. Dordrecht: Kluwer.

Bortolotti, L. 2003: Inconsistency and interpretation. *Philosophical Explorations*, VI(2), 109–123.

Breen, N. *et al.* 2000: Towards an understanding of delusions of misidentification: four case studies. In M. Coltheart and M. Davies (eds.) *Pathologies of Belief.* Oxford: Blackwell, 75–110.

Campbell, J. 2002: Rationality, meaning and the analysis of delusion. *Philosophy, Psychiatry & Psychology*, 8 (2/3), 89–100.

Cherniak, C. 1986: *Minimal Rationality*. Cambridge (Mass.): MIT Press.

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

Davidson, D. 1974: On the very idea of a conceptual scheme. In *Inquiries into Truth and Interpretation*. Oxford: Clarendon Press, 1984.

Davidson, D. 1985a: Incoherence and irrationality. *Dialectica*, 39, 345–354.

Davidson, D. 1985b: Deception and division. In J. Elster (ed.), *The Multiple Self.* Cambridge: Cambridge University Press.

Davies, M. and Coltheart, M. 2000: Introduction. In M. Coltheart and M. Davies (eds.), *Pathologies of Belief.* Oxford: Blackwell, 1–46.

Dennett, D. 1971: Intentional systems. In *Brainstorms*. Cambridge (Mass.): MIT Press.

Emmons, S. *et al*. 1997: *Living with Schizophrenia*. London: Taylor and Francis.

Freud, S. 1917: *Delusion and Dream*. New York: Moffat Yard.

Gerrans, P. 2000: Refining the explanation of the Cotard's delusion. In M. Coltheart and M. Davies (eds.), *Pathologies of Belief*. Oxford: Blackwell, 111–122.

Heal, J. 1998: Understanding other minds from the inside. In A. O'Hear (ed.), *Current Issues in Philosophy of Mind*. Cambridge: Cambridge University Press.

Langdon, R. and Coltheart, M. 2000: The cognitive neuropsychology of delusions. In M. Coltheart and M. Davies (eds.), *Pathologies of Belief*. Oxford: Blackwell, 183–216.

Leeser, J. and O'Donohue, W. 1999: What is a delusion? Epistemological dimensions. *Journal of Abnormal Psychology*, 108 (4), 687–694.

Lukes, S. 1982: Relativism in its place. In M. Hollis and S. Lukes (eds), *Rationality and Relativism*. Oxford: Blackwell.

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

Payne, R.L. 1992: First person account: My Schizophrenia, Schizophrenia Bulletin, 18 (4), 725–728.

Phillips, M.L. and David, A.S. 1997: Visual scan paths are abnormal in deluded schizophrenics. *Neuropyschologia*, 35 (1), 99–105.

Reid, I. *et al*. 1993: Voice recognition impairment in a blind Capgras patient. *Behavioural Neurology*, 6 (4), 225–228.

Sass, L.A. 1994: *The Paradox of Delusion*. Ithaca and London: Cornell University Press.

Silva, J. *et al*. 1994: Delusional misidentification syndromes and dangerousness. *Psychopathology*, 27, 215–219.

Stich, S. 1981: Dennett on intentional systems. *Philosophical Topics*, 12, 39–62.

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

Young, A.W. 2000: Wondrous strange: The neuropsychology of abnormal beliefs. In M. Coltheart and M. Davies (eds.), *Pathologies of Belief*. Oxford: Blackwell, 47–73.