

## **Delusion and Motivationally Biased Belief: Self-Deception in the Two-Factor Framework\***

MARTIN DAVIES

---

**Abstract:** Mele's reflections on the question 'how delusional beliefs are related to self-deception' offer important insights, both about the two-factor framework for understanding delusions and about self-deception. In the two-factor framework, we need to consider, not only the first and second factors, but also the route from experience to belief. We can consider the various points at which motivation might enter the aetiology of a delusion and ask whether the role played by motivation would be enough to produce an example of self-deception. In this exercise, self-deception is revealed as a demanding notion. There are many ways short of self-deception that motivation may play a role in a subject's arriving at a false belief. The cases of delusion that are most clearly examples of self-deception according to Mele's account are those in which motivational bias makes a substantial contribution to the second factor in the aetiology of the delusion.

---

\* I am grateful to Anne Aimola Davies and Tim Bayne for comments on, and conversations about, earlier versions of this chapter.

If motivation is sometimes a factor in the aetiology of delusions, are some cases of delusion also examples of self-deception? Alfred Mele's reflections on this question offer important insights, both about the two-factor framework for understanding delusions and about self-deception. The lesson about the two-factor framework is this. In order to consider systematically the points at which motivation might enter the aetiology of a delusion, we need to consider, not only the first factor and the second factor, but also the processing stage that leads from experience to belief. The lesson about self-deception is that it is a demanding notion. There are many ways short of self-deception that motivation may play a role in a subject's arriving at a false belief.

I begin with the two-factor framework and a brief review of Mele's account of self-deception. Then I consider the various points at which motivation might enter the aetiology of a delusion and ask, in each case, whether the role played by motivation would be enough to produce an example of self-deception.

## **1. The Two-Factor Framework: The Anomalous Experience Hypothesis**

Brendan Maher (1974, 1988, 1992, 1999, 2003) says that delusions are false beliefs that arise as normal responses to anomalous experiences. According to Maher's *anomalous experience hypothesis*, if delusions are pathologies of belief then the locus of pathology lies in experience and not, for example, in the subject's reasoning. According to the two-factor framework for understanding delusions, however, there must be a second locus of pathology in order to account for the fact that the false belief is not just adopted but is 'firmly sustained despite what almost everyone else believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary' (*DSM-IV-TR*, 2000, p. 821). Even if an anomalous experience provides an answer, or the beginnings of an answer, to the question, 'Where did the delusion come from?', it does not provide any answer to the question, 'Why does the patient not reject the belief?' (Coltheart, 2007, p. 1044).

### *1.1 Feelings of significance*

Maher is primarily concerned with the role of anomalous experiences in the aetiology of delusions in schizophrenia and he proposes that the subject of a schizophrenic delusion experiences unusual *feelings of significance*.

In a famous example 'a patient . . . looked at the marble tables in a café and suddenly became convinced that the end of the world was coming' (Maher, 1999, p. 559). Maher's proposal about this case is not that the patient had an anomalous visual experience of marble tables but that the patient experienced feelings of heightened significance when viewing the marble tables. From the patient's point of view, such feelings of significance cry out for explanation in terms of something that has changed and 'if no specific concrete change can be found, and the feeling of significance persists, everything must have changed in some fundamental way' (*ibid.*, p. 560). The patient may seek to explain a persistent and non-specific feeling of significance in apocalyptic terms.

Feelings of significance are, of course, common in everyday life (*ibid.*, p. 558):

Survival requires the existence of a detector of changes in the normally regular patterns of environmental stimuli, namely those that are typically dealt with automatically. The detector functions as a general non-specific alarm, a

‘significance generator’, which then alerts the individual to scan the environment to find out what has changed.

We sometimes have a vague feeling that something has changed or that something is not quite right. This feeling normally prompts a search for a specific difference in the local environment and sometimes the search leads us to a substantially incorrect view about *what* has changed. Maher himself describes a familiar kind of example (ibid., p. 553):

This writer once shaved off his beard and the next day was greeted with comments ranging from direct recognition of what had changed, to the case of one colleague who commented, ‘There’s something different about you. Oh yes. You are wearing a different kind of tie.’

### *1.2 Anomalous experience in the Capgras delusion*

Max Coltheart (2005) appeals to something like Maher’s ‘detector of changes’ or ‘significance generator’ in an answer to the question where the Capgras delusion comes from – an answer that builds on Hadyn Ellis and Andrew Young’s (1990, p. 244) suggestion that the Capgras delusion is the ‘mirror image’ of prosopagnosia with autonomic signs of covert recognition.

Some patients in whom overt face recognition is impaired (patients with prosopagnosia) nevertheless show autonomic responses that distinguish familiar from unfamiliar faces (Tranel and Damasio, 1985). Ellis and Young propose that the reverse dissociation of impairments is found in Capgras patients: overt face recognition is intact but the normal autonomic response to a familiar face (such as the face of a spouse) is absent. Ellis and colleagues (Ellis, Young, Quayle and de Pauw, 1997), William Hirstein and V.S. Ramachandran (1997), and Gianni Brighetti and colleagues (Brighetti, Bonifacci, Borlimi and Ottaviani, 2007) provide evidence supporting this proposal. In patients with the Capgras delusion, skin conductance response (a measure of autonomic response) does not distinguish between familiar faces (famous faces or family faces) and unfamiliar faces. Ellis and Young (1990) thus suggest that the Capgras patient has an experience of seeing a face that looks just like the spouse, but without the affective response that would normally be an integral part of that experience. The delusion then arises as the patient tries to explain this peculiar experience.

This suggestion as to where the Capgras delusion comes from requires some further development because there is no reason to suppose that the presence or absence of activity in the autonomic nervous system figures, as such, in a person’s conscious experience. For example, patients who have prosopagnosia with covert recognition do not seem to experience their intact autonomic responses to familiar faces (Coltheart, Langdon and Breen, 1997).

In order to avoid commitment to the idea that the Capgras patient *experiences* the absence of autonomic activity as such, Coltheart appeals to two automatic and *unconscious* processes. One is a process of prediction (2005, p. 155):

It is a general principle of cognitive life that we are continually making predictions, on the basis of what we currently know about the world, concerning what will happen to us next.

The other is a process of comparison between what is predicted and what actually happens (ibid.):

Only when a prediction fails does consciousness get involved; the unconscious system makes some kind of report to consciousness to instigate some intelligent conscious problem-solving behaviour that will discover what's wrong with the database [used to generate the predictions] and how it should be repaired.

Coltheart's proposal that an unconscious comparator system 'makes some kind of report to consciousness' leading to 'behaviour that will discover what's wrong' is evidently very similar to Maher's proposal that a detector of changes 'alerts the individual to scan the environment to find out what has changed' (1999, p. 558).

In the case of a patient who has the Capgras delusion, believing that his wife has been replaced by an impostor, Coltheart describes the operation of the subpersonal-level comparator system this way (2005, p. 155): 'the unconscious system predict[s] that when the wife is next seen a high autonomic response will occur, detect[s] that this does not occur, and report[s] to consciousness, "There's something odd about this woman"'. The patient then engages in 'intelligent conscious problem-solving behaviour' directed towards explaining this sense of oddity, change, or significance. The patient's belief that this woman is not, in fact, his wife arises within this personal-level project. Coltheart thus gives an *explanationist* answer to the question where the Capgras delusion comes from.

## **2. The Two-Factor Framework: From Anomalous Experience to Delusional Belief**

The idea that the route from anomalous experience to delusional belief is one in which the subject tries to *explain* the experience is already present in Maher (1974, p. 103): 'a delusion is a hypothesis designed to explain unusual perceptual phenomena'. Tony Stone and Andrew Young say that 'the Capgras delusion is the sufferer's attempt to *account for* an anomalous perceptual experience' (1997, p. 338; emphasis added). Peter Halligan and colleagues speak of the patient's '*interpretation* of what he discovers has . . . happened to him' and of the patient as 'trying to *understand* the (abnormal) experiences . . . that are consequent upon the brain damage he has sustained' (Halligan, Marshall and Wade, 1995, p. 179).

### *2.1 Alternatives to the explanationist answer*

The explanationist answer to the question where the Capgras delusion comes from can be contrasted with two others.

According to one alternative answer, the representational content of the Capgras patient's experience is more specific than 'This is someone who looks just like my wife *but there is something odd about her*'. It is, rather: 'This is someone who looks just like my wife *but it is not really her*.' Here, the suggestion is that, in the absence of the normal autonomic response, the face processing system delivers an anomalous experience that encodes the core content of the delusion, or something very close to it. There is scarcely daylight between the proposition that this person who looks like my wife (and, indeed, claims to be my wife) is not really my wife and the proposition that this person is an impostor.

If the content of the delusion is already encoded in experience then the personal-level step from the experience to the delusion is not one of explaining an anomalous experience but simply one of taking the experience to be veridical (Davies, Coltheart, Langdon and Breen, 2001). Tim Bayne and Elisabeth Pacherie (2004) describe this kind of case in terms of the patient *endorsing* the experience; Cordelia Fine and colleagues say that the patient's belief is an *expression* of the content of the experience (Fine, Craigie and Gold, 2005; see also Gold and Hohwy, 2000).

A second alternative answer to the question where the Capgras delusion comes from offers a more dramatic departure from Maher's anomalous experience hypothesis. According to this answer, anomalous *experience* does not figure in the aetiology of the delusion at all. The route from the absence of a normal response in the patient's autonomic nervous system to the patient's considering the hypothesis or initially adopting the belief that his wife has been replaced by an impostor lies wholly at the subpersonal level and involves wholly unconscious processes. The patient simply finds himself with this idea, hypothesis, or belief. It may be that the content of the delusion seems to be present in the patient's experience when he sees his wife; but if this is so then it is the result of a process that John Campbell describes as 'top-down loading of the perception by the delusional content' (2001, p. 96).<sup>1</sup>

## 2.2 Routes from deficit to delusion: Personal or subpersonal

Maher says (1999, p. 551):

The origins of anomalous experience may lie in a broad band of neuropsychological anomalies. These include, but are not confined to . . . endogenous neural activation of the feeling of significance normally triggered by pre-conscious recognition of changes in a familiar environment.

In the case of the Capgras delusion, the suggested neuropsychological anomaly is disruption of the connection between the patient's face recognition system and autonomic nervous system. This neuropsychological anomaly or deficit figures in the answer to the question, 'Where did the delusion come from?' But there is more than one possible route from the neuropsychological deficit to the initial adoption of the false belief.

In general, the route from deficit to delusion might lie mainly at the personal level or mainly at the subpersonal level. According to Maher's account, the neuropsychological deficit gives rise to an anomalous experience with a relatively non-specific content and personal-level processes of explanation and interpretation lead from that experience to the false belief. According to the account that departs dramatically from Maher's, personal-level processes have no role to play and the route from deficit to delusion lies wholly at the subpersonal level. According to the account that is intermediate between these extremes, the neuropsychological deficit gives rise to an anomalous experience that encodes the content of the delusion so that only the simple personal-level process of endorsement is required.

---

<sup>1</sup> Campbell continues (2001, p. 96): 'this loading of the perception may indeed be instrumental in the maintenance of the delusion, but . . . the top-down loading cannot be regarded as itself the source of the delusion'.

My aim is to explore the relationship between delusion and self-deception by considering points at which motivational bias – intuitively, a personal-level phenomenon – might enter the aetiology of delusions. Consequently, in later sections I shall focus on the two mainly personal-level routes from deficit to delusion, the route via endorsement of an anomalous experience and the route via explanation of an anomalous experience. Before that, I offer a brief review of Mele’s account of self-deception.

### 3. Mele’s Deflationary Account of Self-Deception

One way to start thinking about self-deception is to consider it as structurally analogous to intentional deception of another person. We are then led naturally to the doctrine that a self-deceived subject starts out believing a true proposition not-P and intentionally acts to bring it about that he or she (himself or herself) ends up believing the false proposition P. This doctrine about self-deception seems problematic because it presents two puzzles. First, according to the doctrine, the self-deceived subject will often, perhaps typically, end up believing both the true proposition not-P and the false proposition P. This is what Mele (2001, p. 7) calls the *static* puzzle about self-deception. Second, there is also a *dynamic* puzzle (ibid., p. 8) concerning the process of deception in which, according to the doctrine, the subject intentionally, and in typical cases knowledgeably, engages.

Mele (1997, 2001) has developed and defended an account of self-deception that avoids commitment to the apparently problematic doctrine and is ‘both more subtle and less problematic than interpersonal models imply’ (2001, p. 25).<sup>2</sup>

#### 3.1 The basic idea

The basic idea of Mele’s deflationary account of self-deception is that, as a result of *motivationally biased handling of the evidence*, a subject, S, ends up believing the false proposition P even though the evidence predominantly supports the true proposition not-P. Mele is explicit that he is not offering a conceptual analysis of self-deception. But he elucidates the notion by proposing a set of four jointly *sufficient* conditions for self-deception (2001, pp. 50–1).

One of the four conditions proposed as jointly sufficient for self-deception says that ‘the body of data possessed by S at the time provides greater warrant for not-P than for P’ (2001, p. 51, condition 4). Some philosophers say that this is also a necessary condition. But Mele suggests that believing against the weight of the evidence that one *possesses* is not essential for self-deception. Suppose that the evidence that is *readily available* to the subject, S, strongly favours the true proposition not-P. Suppose, too, that S *gathers* evidence in such a motivationally biased way that the evidence that she actually possesses favours the false proposition P rather than the true proposition not-P. If, as the result of her motivationally biased evidence-gathering, S ends up believing P, then she is ‘naturally deemed self-deceived, other things being equal’ (ibid., p. 52).

---

<sup>2</sup> See Levy (this volume) for the proposal that cases of anosognosia show that the apparently problematic doctrine really does apply to at least some examples of self-deception.

### 3.2 *The impartial observer test*

In line with the basic idea of his account, Mele proposes the *impartial observer test* (2001, p. 106; this volume, p. 000) as a test for self-deception – or, more accurately, ‘a test for the satisfaction of a *necessary* condition for self-deception’ (2001, p. 121; emphasis added).

Suppose that we are considering whether a subject S is self-deceived in believing the false proposition P. The basic idea says that S is self-deceived if she believes P as a result of motivationally biased handling of a body of available evidence that predominantly supports the true proposition not-P. The impartial observer test says (2001, p. 106):

If S is self-deceived in believing P, and D is the collection of relevant data readily available to S, then if D were made readily available to S’s impartial cognitive peers . . . those who conclude that P is false would significantly outnumber those who conclude that P is true.

For the purpose of this test, the *impartial cognitive peers* of the subject S are actual or hypothetical subjects who are similar to S in their cognitive capacities and who, in point of their desires and emotions, have nothing at stake in the question whether P or not-P is true. They do not, for example, prefer the truth of P over the truth of not-P, nor *vice versa*. They do not regard believing P in error as preferable over believing not-P in error, nor *vice versa*.

Passing the impartial observer test is only a necessary condition for self-deception and it is intuitively clear how a subject might pass the test without being self-deceived in believing P. The basic idea of Mele’s account is that self-deception involves motivationally biased *handling of the available evidence*. But there are other ways in which a preference for the truth of P over the truth of not-P, or a preference for believing P in error over believing not-P in error, might bias a subject’s belief. Motivation can, for example, bias the selection of hypotheses and such a motivationally biased selection can have consequences for the operation of further processes that are not themselves motivationally biased. Mele gives the example of motivated selection of hypotheses ‘setting the stage for the confirmation bias’ (2001, p. 30). Assume for the moment that the bias in favour of evidence that confirms the hypothesis under consideration is an unmotivated or ‘cold’ bias. Then it is plausible that the subject arrives at a motivationally biased belief but is not self-deceived.

## 4. **Motivational Bias in the Aetiology of Delusions**

We can now consider the various points at which motivation might enter the aetiology of a delusion, in the first or second factor, or along the route from experience to belief. We shall ask, in each case, whether the role played by motivation would be enough to produce an example of self-deception. As a baseline for comparison with subsequent cases, consider first a hypothetical case of delusion in which motivation does not figure at all.

### 4.1 *Case 0: Delusion without motivation*

A subject, S<sub>0</sub>, has adopted a false proposition P as a belief. We stipulate that the answer to the question, ‘Where did the delusion come from?’, does not involve motivation.

Wherever the false belief came from, it is a delusion because it is maintained despite a mass of available evidence that speaks against it. We also stipulate that motivation does not figure in the answer to the question, ‘Why does the subject not reject the belief?’. The subject, S0, maintains the false belief P as the result of impairment to a cognitive system that normally functions ‘to evaluate beliefs with respect to their plausibility and the evidence for and against them and to accept or reject beliefs on this basis’ (Coltheart, 2007, p. 1047).

This hypothetical case of delusion (Case 0) is not an example of self-deception according to the basic idea of Mele’s account. There is no motivationally biased handling of the available evidence. The case also appears not to pass the impartial observer test. Motivation does not enter the account of how S0 first came to adopt P as a belief. So the impartial cognitive peers of S0 would (or might) initially adopt P as a belief, just as S0 did. And then – because, as cognitive peers, they share S0’s cognitive impairment – they would (or might) fail to reject the proposition P, despite the mass of available evidence against it. The impartial cognitive peers who conclude that P is false and reject the initially adopted belief would not significantly outnumber those who conclude that P is true and maintain the belief. In cases of this kind, ‘the delusion seems to lie well beyond the sphere of self-deception’ (Mele, this volume, p. 000).

Mele suggests that many cases of the Capgras delusion are similar to Case 0. He considers the case of Carl who, as the result of a neuropsychological deficit, ‘has a new way of experiencing his wife’s face’ (this volume, p. 000). Carl initially adopts the false belief that his wife has been replaced by an impostor and then fails to reject this belief because he has an impairment of the cognitive system that normally functions to evaluate beliefs.

In order to apply the impartial observer test, we need to consider those of Carl’s actual or hypothetical impartial cognitive peers to whom the same evidence is available. Mele suggests that we ‘limit members of the panel . . . to people who have experiences of the sort characteristic of Capgras patients’ (p. 000). I assume that these cognitive peers have a new way of experiencing their own spouse’s face, rather than the face of Carl’s wife and that they would (or might) initially adopt the belief that their own spouse has been replaced by an impostor. They may well regard *Carl’s* belief about *his* wife as false and absurd. They may think that Carl has gone mad (Young, 1998, p. 37). But because, as cognitive peers, they share Carl’s cognitive impairment, many or most of them will fail to reject their false belief about their own spouse, even though they may appreciate that other people will regard their belief as implausible.

#### *4.2 Case 1: Motivational bias in the first factor*

Suppose (perhaps *per improbable*) that a subject, S1, undergoes an anomalous experience that is produced by a motivational bias acting alone or in concert with some neuropsychological deficit. As the result of this experience, S1 initially adopts a false proposition P as a belief and no additional motivational factors are involved in the answer to the question, ‘Where did the delusion come from?’. As the result of a second factor that is neuropsychological, rather than motivational, in nature, S1 maintains this false belief despite a mass of available evidence that speaks against it. This hypothetical case of delusion (Case 1) is not an example of self-deception according to the basic idea of

Mele's account. Motivational bias figures in the first factor by generating a piece of relevant evidence – namely, the anomalous experience. But the available evidence is not handled in a motivationally biased way.

The case also appears not to pass the impartial observer test. At first, it might be thought that the impartial cognitive peers of S1 would not share S1's anomalous experience and so would not initially adopt the false proposition as a belief. But, in order to apply the test, we have to consider impartial cognitive peers to whom the same evidence is available. So we must stipulate that the impartial cognitive peers *would* share S1's anomalous experience. Consequently, they would (or might) adopt and maintain the false proposition P as a belief, just as S1 did. The impartial cognitive peers who conclude that P is false would not significantly outnumber those who conclude that P is true.

Mele (this volume) tentatively suggests a way in which some cases of the Capgras delusion and the Cotard delusion might be similar to Case 1. Mele's suggestion draws on Young's proposal that the Capgras patient and the Cotard patient might have similar anomalous experiences but explain their experiences in different ways, reflecting differences in mood and attributional style. The Capgras patient locates the cause of his anomalous experience in the external world – 'My wife has been replaced by an impostor' – while the Cotard patient locates the cause within himself – 'I am dead' (Wright, Young and Hellawell, 1993; Young, 1999, 2000).

Mele's variation on this theme involves the suggestion that, rather than *explaining* similar anomalous experiences in different ways, the patients *endorse* different anomalous experiences that arise from similar neuropsychological deficits. That is, Mele suggests that the Capgras patient and the Cotard patient might have similar neuropsychological deficits that act in concert with different affective and motivational factors to produce different experiences that encode the core contents of their respective delusions. These factors would figure in the aetiology of the delusions but 'they would do so by helping to cause experiential data . . . rather than by biasing the person's treatment of data' (this volume, p. 000).

#### *4.3 Case 2: Motivational bias in the second factor*

A subject, S2, has initially adopted a false proposition P as a belief and the belief is sustained despite a mass of available evidence that speaks against P and in favour of the true proposition not-P. As in Case 0, we stipulate that the answer to the question, 'Where did the delusion come from?', does not involve motivation. But why does S2 not reject the false belief? Let us suppose, as the generic form of the two-factor framework surely allows, that S2 maintains the false belief P as the result of motivationally biased treatment of the available evidence. The subject S2's preference for the truth of P over the truth of not-P is manifested in biased gathering of evidence, biased attention to evidence, and biased interpretation of evidence (Mele, 2001, pp. 26–7). Thus, the hypothetical case of delusion (Case 2) is also an example of self-deception according to the basic idea of Mele's account.

The case also appears to pass the impartial observer test. We stipulated that motivation does not enter the account of how S2 first came to adopt P as a belief. So the impartial cognitive peers of S2 would (or might) initially adopt P as a belief, just as S2 did. But then, being impartial rather than motivationally biased in their handling of the

available evidence, they would subsequently reject the proposition P on the basis of the mass of available evidence against it. The impartial cognitive peers who conclude that P is false would significantly outnumber those who conclude that P is true.

Mele (this volume) suggests that Peter Butler's (2000) case of reverse Othello syndrome, patient BX, may be similar to Case 2. Whatever the answer to the question where BX's delusion came from, motivationally biased treatment of the available evidence – 'selective focusing or attending' – may have figured as the second factor (this volume, p. 000): 'Even if he [BX] does not enter self-deception in acquiring the belief that he is married to N [BX's former partner, who separated from him after his injury], he may be self-deceived in continuing to believe this.'

In short, cases of delusion in which motivationally biased treatment of the available evidence makes a substantial contribution to the second factor will plausibly be examples of self-deception according to the basic idea of Mele's account and will pass the impartial observer test. Specifically, they will be examples in which the subject is self-deceived in *maintaining* the false belief.

#### *4.4 Case 3: Motivational bias on the route from experience to belief*

In his discussion of delusions, Mele mainly adopts (at least for expository purposes) the endorsement account of the route from anomalous experience to false belief. According to the endorsement account, the route involves just one short step of taking the experience to be veridical and there is little space for motivationally biased handling of evidence before the initial adoption of the false belief. Handling of the available evidence – whether the handling is biased, impaired, or normal – is part of a subsequent process of belief evaluation. We should, however, consider a hypothetical case of delusion in which motivationally biased handling of the available evidence does play a role before the subject's initial adoption of the false belief, and so earlier than the second factor (Case 3).

A subject, S3, has an anomalous experience and, in attempting to explain, understand, or interpret it, is led to consider a hypothesis, P, that is actually false. The hypothesis is unattractive, and S3 hopes that it is not true; but S3 would very much prefer to make the error of believing P when not-P was really true rather than the error of believing not-P when P was really true. The costs to S3 of believing not-P in error are high and so S3's handling of the available evidence is motivationally biased. The evidence speaks in favour of not-P, but S3 comes to believe P. Suppose that, having initially adopted this false belief, S3 subsequently evaluates it. We can add to the case that, because of the operation of the same biases, the false belief is maintained.

Case 3 is clearly an example of self-deception according to the basic idea of Mele's account. The subject, S3, is self-deceived in *adopting* the false proposition P as a belief. Furthermore, if the same motivationally biased handling of the available evidence is at work in subsequent belief evaluation then S3 is self-deceived in *maintaining* the belief. Case 3 also passes the impartial observer test. The impartial cognitive peers of S3, having nothing at stake in the question whether P or not-P is true, would reject the proposition P on the basis of the mass of evidence against it.

When motivationally biased handling of the evidence is a factor in the adoption of a false belief, it is not guaranteed that the same biases will be operative in subsequent belief evaluation. In principle, S3 could be self-deceived in initially adopting the false belief P,

even though it was a cognitive impairment that explained why the false belief, once adopted, was maintained rather than rejected. But a mixed case of that kind raises the question whether a cognitive impairment that would prevent belief evaluation could still allow a subject to engage in motivationally biased handling of the evidence in initially arriving at a belief. It may well be that often, if biased handling of the evidence figures on the route from experience to belief, it also figures in the second factor.

Mele (this volume) shows that some cases of delusional jealousy are similar to Case 3. They are examples of self-deception in virtue of motivational bias operating before the subject's initial adoption of the false belief that his or her partner is unfaithful. A subject who would much prefer the error of believing a faithful partner to be unfaithful over the error of believing an unfaithful partner to be faithful could be motivationally biased towards gathering and attending to evidence of infidelity and towards interpreting evidence as supporting the hypothesis of infidelity and could have 'an extremely high threshold for accepting the fidelity proposition' (this volume, p. 000). Such a subject would count as self-deceived in initially adopting the false belief in the partner's infidelity. If the same biases were operative in subsequent belief evaluation then the subject could also count as self-deceived in maintaining, instead of rejecting, the false belief.

## **5. Motivation Without Self-Deception**

Mele suggests that his account of delusional jealousy as self-deception may be especially persuasive in cases where delusional jealousy is the subject's only delusion. He allows that it may be less convincing when other delusions are present if we are 'inclined to see [these other delusions] as falling outside the sphere of self-deception and as being explained in part by a cognitive deficit' (this volume, p. 000). In this connection, he specifically mentions persecutory delusions but, in fact, theories of persecutory delusions often appeal to motivational factors (see Bentall, Corcoran, Howard, Blackwood and Kinderman, 2001, for a review).

### *5.1 Motivational factors in persecutory delusions*

Richard Bentall and colleagues propose that patients with persecutory delusions have an abnormal attributional style, including 'a tendency to attribute negative events to powerful others' (2001, p. 1158). They suggest that the externalising attributional style functions to preserve self-esteem and this *persecutory delusions as defence* hypothesis yields the prediction that individuals with persecutory delusions will show discrepancies between overt and covert self-esteem. However, while the evidence for an externalising bias in patients with persecutory delusions appears to be strong, the evidence on the prediction about overt and covert self-esteem is mixed and so 'the "persecutory delusion as defence" hypothesis is less well supported' (Blackwood, Howard, Bentall and Murray, 2001, p. 528).

Another motivational construct that may be a factor in persecutory delusions is *need for closure*, defined by Arie Kruglanski as 'the desire for a definite answer on some topic, any answer compared to confusion and ambiguity' (Kruglanski, 1989, p. 14; see also Kruglanski, Webster and Klem, 1993). Susannah Colbert and Emmanuelle Peters (2002) found higher-than-normal need for closure in a non-clinical group of delusion-prone

individuals and proposed that need for closure may drive the ‘jumping to conclusions’ bias that is found in some patients with delusions (for reviews, see Garety and Freeman, 1999; Fine, Gardner, Craigie and Gold, 2007). In line with this proposal, Bentall and colleagues suggest that the ‘jumping to conclusions’ bias reflects ‘motivational rather than information processing abnormalities’ (Bentall et al., 2001, p. 1162). Again, however, recent evidence has been mixed and, in a large cross-sectional study, Daniel Freeman and colleagues (Freeman, Garety, Kuipers, Colbert, Jolley, Fowler, Dunn and Bebbington, 2006) found that need for closure scores were not related to jumping to conclusions and concluded, more generally, that there is no evidence that need for closure (as assessed by Kruglanski’s Need for Closure Scale) ‘could be a proximal cause for delusions’ (Freeman et al., 2006, p. 1147).

While the evidence supporting these two proposals about motivational factors in persecutory delusions is far from compelling, they do illustrate two general points about self-deception. First, need for closure is a motivational factor and, if Colbert and Peters (2002) were right, it would drive a data-gathering bias, a bias towards early acceptance of hypotheses. But this motivationally driven bias in handling evidence would not be a motivational bias in the sense that is required by Mele’s account of self-deception. Need for closure is not a preference for the truth of P over the truth of not-P, nor a preference for believing P in error over believing not-P in error. It is an impartial preference for any answer, whether P or not-P, over confusion and ambiguity.

Second, whether or not an externalising attributional style is motivated by a need to preserve self-esteem, it does seem to involve a preference for the truth of some hypotheses over others. Individuals with this attributional style prefer to assign causal responsibility for negative events externally rather than internally. Attributional bias is a fine example in support of Mele’s claim that ‘motivation can influence which hypotheses occur to one and affect the salience of available hypotheses’ (2001, pp. 29–30). But (as we noted towards the end of section 3), motivationally biased hypothesis selection is not yet motivationally biased handling of evidence.

#### *5.2 Case 4: Motivational bias on the route from experience to belief – again*

We can use these two points in a final hypothetical case of delusion where we assume an explanationist account of the route from anomalous experience to false belief.

A subject, S4, has a neuropsychological deficit, giving rise to an anomalous experience that cries out for explanation. Motivationally biased processes lead to the selection of the false proposition P as a hypothesis that is putatively explanatory of that experience. To sharpen the case, we can stipulate that it is exceedingly unlikely that any subject without S4’s motivational bias would consider the hypothesis P to be at all credible. A bias towards early acceptance of hypotheses then leads S4 to adopt the false proposition P as a belief on the basis of very little evidence. This data-gathering bias is produced either by a cognitive impairment or else by an impartial motivational factor like need for closure. The second factor in the aetiology of S4’s delusion is a neuropsychological deficit in the cognitive processes of belief evaluation. This deficit has the consequence that the false belief is not rejected even though the available evidence tells massively against it.

The subject S4's false belief P is motivationally biased to the extent that S4's hypothesis selection is motivationally biased; but that does not settle whether the hypothetical case of delusion (Case 4) is an example of self-deception according to Mele's account. The argument for saying that it is not an example of self-deception is simply that, as described, it does not involve motivationally biased handling of the available evidence. In the processing stages leading from anomalous experience to false belief, there is very little handling of the available evidence at all; and even in the version where there is a motivational factor at work, it is an impartial factor and not a motivational bias in the sense required by Mele's account. In the processing stages following the initial adoption of the false belief, it is maintained rather than rejected, not because of motivationally biased handling of the evidence, but because of cognitive impairment.

Case 4 does, however, appear to pass the impartial observer test. The impartial cognitive peers of S4 would share S4's anomalous experience, evidence-gathering bias, and impaired belief evaluation, but they would differ from S4 in a crucial way. They would not share S4's motivationally biased hypothesis selection and we have stipulated that it is unlikely that any subject without that bias would consider the hypothesis P to be at all credible. If the proposition P were to achieve some credibility and be initially adopted as a belief then cognitively impaired subjects would be unable to reject it, just as S4 was unable to reject it. But the impartial cognitive peers of S4 would reject the hypothesis without needing to rely on their impaired cognitive processes of belief evaluation. The cognitive peers who conclude that P is false would significantly outnumber those who conclude that P is true.

It is, of course, no objection to Mele's overall account that Case 4 is not an example of self-deception according to the basic idea of his account yet passes the impartial observer test. Passing the test is only a necessary condition for self-deception.

## **Conclusion**

According to the two-factor framework for understanding delusions, motivational factors might enter the aetiology of a delusion in the first or second factor, or along the route from experience to belief. According to Mele's account of self-deception, cases of delusion in which motivational or affective factors figure in the first factor as causes of an anomalous experience are not examples of self-deception (Case 1). In contrast, cases of delusion where motivational bias figures in the second factor are likely to be examples of self-deception (Case 2). Some cases of delusion in which motivational factors operate earlier than the second factor are also examples of self-deception (Case 3). But many cases in which motivation enters the processing stages leading up to initial adoption of the false belief are not examples of self-deception because they do not involve motivationally biased handling of the available evidence (Case 4).

Self-deception is a demanding notion. Overall, the cases of delusion that are most clearly examples of self-deception according to Mele's account are those in which motivational bias makes a substantial contribution to the second factor in the aetiology of the delusion.

## References

- American Psychiatric Association 2000: *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR)*. Washington, DC: American Psychiatric Association.
- Bayne, T. and Pacherie, E. 2004: Bottom-up or topdown? Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry, and Psychology*, 11, 1–11.
- Bentall, R.P., Corcoran, R., Howard, R., Blackwood, N. and Kinderman, P. 2001: Persecutory delusions: A review and theoretical integration. *Clinical Psychology Review*, 21, 1143–92.
- Blackwood, N.J., Howard, R.J., Bentall, R.P. and Murray, R.M. 2001: Cognitive neuropsychiatric models of persecutory delusions. *American Journal of Psychiatry*, 158, 527–39.
- Brighetti, G., Bonifacci, P., Borlimi, R. and Ottaviani, C. 2007: 'Far from the heart far from the eye': Evidence from the Capgras delusion. *Cognitive Neuropsychiatry*, 12, 189–97.
- Butler, P.V. 2000: Reverse Othello syndrome subsequent to traumatic brain injury. *Psychiatry*, 63, 85–92.
- Campbell, J. 2001: Rationality, meaning, and the analysis of delusion. *Philosophy, Psychiatry and Psychology*, 8, 89–100.
- Colbert, S.M. and Peters, E.R. 2002: Need for closure and jumping-to-conclusions in delusion-prone individuals. *Journal of Nervous and Mental Disease*, 190, 27–31.
- Coltheart, M. 2005: Conscious experience and delusional belief. *Philosophy, Psychiatry, and Psychology*, 12, 153–7.
- Coltheart, M. 2007: Cognitive neuropsychiatry and delusional belief. *Quarterly Journal of Experimental Psychology*, 60, 1041–62.
- Coltheart, M., Langdon, R. and Breen, N. 1997: Misidentification syndromes and cognitive neuropsychiatry. *Trends in Cognitive Sciences*, 1, 157–8.
- Davies, M., Coltheart, M., Langdon, R. and Breen, N. 2001: Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, and Psychology*, 8, 133–58.
- Ellis, H. D. and Young, A. W. 1990: Accounting for delusional misidentifications. *British Journal of Psychiatry*, 157, 239–48.
- Ellis, H.D., Young, A.W., Quayle, A.H. and de Pauw, K.W. 1997: Reduced autonomic responses to faces in Capgras delusion. *Proceedings of the Royal Society: Biological Sciences*, B264, 1085–92.
- Fine, C., Craigie, J. and Gold, I. 2005: Damned if you do, damned if you don't: The impasse in cognitive accounts of the Capgras delusion. *Philosophy, Psychiatry, and Psychology*, 12, 143–51.
- Fine, C., Gardner, M., Craigie, J. and Gold, I. 2007: Hopping, skipping or jumping to conclusions? Clarifying the role of the JTC bias in delusions. *Cognitive Neuropsychiatry*, 12, 46–77.
- Freeman, D., Garety, P.A., Kuipers, E., Colbert, S., Jolley, S., Fowler, D., Dunn, G. and Bebbington, P.A. 2006: Delusions and decision-making style: Use of the Need for Closure Scale. *Behaviour Research and Therapy*, 44, 1147–58.

- Garety, P.A. and Freeman, D. 1999: Cognitive approaches to delusions: A critical review of theories and evidence. *British Journal of Clinical Psychology*, 38, 113–54.
- Gold, I. and Hohwy, J. 2000: Rationality and schizophrenic delusion. *Mind and Language*, 15, 146–67.
- Halligan, P. W., Marshall, J. C. and Wade, D. T. 1995: Unilateral somatoparaphrenia after right hemisphere stroke: A case description. *Cortex*, 31, 173–82.
- Hirstein, W. and Ramachandran, V.S. 1997: Capgras syndrome: A novel probe for understanding the neural representation of the identity and familiarity of persons. *Proceedings of the Royal Society: Biological Sciences*, B264, 437–44.
- Kruglanski, A.W. 1989: *Lay Epistemics and Human Knowledge: Cognitive and Motivational Bases*. New York: Plenum.
- Kruglanski, A.W., Webster, D.M. and Klem, A. 1993: Motivated resistance and openness to persuasion in the presence or absence of prior information. *Journal of Personality and Social Psychology*, 65, 861–76.
- Maher, B.A. 1974: Delusional thinking and perceptual disorder. *Journal of Individual Psychology*, 30, 98–113.
- Maher, B.A. 1988: Anomalous experience and delusional thinking: The logic of explanations. In T.F. Oltmanns and B.A. Maher (eds), *Delusional Beliefs*. Chichester: John Wiley and Sons, pp. 15–33.
- Maher, B.A. 1992: Delusions: Contemporary etiological hypotheses. *Psychiatric Annals*, 22, 260–8.
- Maher, B.A. 1999: Anomalous experience in everyday life: Its significance for psychopathology. *The Monist*, 82, 547–70.
- Maher, B.A. 2003: Schizophrenia, aberrant utterance and delusions of control: The disconnection of speech and thought, and the connection of experience and belief. *Mind and Language*, 18, 1–22.
- Mele, A.R. 1997: Real self-deception. *Behavioral and Brain Sciences*, 20, 91–102.
- Mele, A.R. 2001: *Self-Deception Unmasked*. Princeton, NJ: Princeton University Press.
- Stone, T. and Young, A.W. 1997: Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language*, 12, 327–64.
- Tranel, D. and Damasio, A.R. 1985: Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science*, 228, 1453–4.
- Wright, S., Young, A.W. and Hellawell, D.J. 1993: Sequential Cotard and Capgras delusions. *British Journal of Clinical Psychology*, 32, 345–9.
- Young, A.W. 1998: *Face and Mind*. Oxford: Oxford University Press.
- Young, A.W. 1999: Delusions. *The Monist*, 82, 571–89.
- Young, A.W. 2000: Wondrous strange: The neuropsychology of abnormal beliefs. *Mind and Language*, 15, 47–73.