

# SELF-DECEPTION AND DELUSIONS

ALFRED R. MELE

Florida State University

## ABSTRACT

My central question in this paper is how delusional beliefs are related to self-deception. In section 1, I summarize my position on what self-deception is and how representative instances of it are to be explained. I turn to delusions in section 2, where I focus on the Capgras delusion, delusional jealousy (or the Othello syndrome), and the reverse Othello syndrome.

**Key words:** self-deception, delusion, Capgras delusion, jealousy, Othello syndrome

My central question in this paper is how delusional beliefs are related to self-deception. In section 1, I summarize my position on what self-deception is and how representative instances of it are to be explained. I turn to delusions in section 2.

## 1. Self-deception

According to a traditional view, self-deception is an intrapersonal analogue of stereotypical interpersonal deception.<sup>1</sup> In the latter case, deceivers *intentionally* deceive others into believing something,  $p$ , and there is a time at which the deceivers believe that  $p$  is false while their victims falsely believe that  $p$  is true. If self-deception is properly understood on this model, self-deceivers intentionally deceive themselves into believing something,  $p$ , and there is a time at which they believe that  $p$  is false while also believing that  $p$  is true.

<sup>1</sup> For citations of this tradition in philosophy, psychology, psychiatry, and biology, see Mele 2001, p. 125, n. 1. *Stereotypical* interpersonal deception does not exhaust interpersonal deception.

In Mele 2001, I criticize this view and defend an alternative, deflationary view, according to which self-deception does not entail any of the following: intentionally deceiving oneself; intending (or trying) to deceive oneself, or to make it easier for oneself to believe something; concurrently believing each of two explicitly contradictory propositions. I also argue that, in fact, ordinary instances of self-deception do not include any of these things. Of course, simply falsely believing that  $p$  in the absence of deception by anyone else is not sufficient for self-deception. If it were, we would be self-deceived whenever we make, for example, unmotivated arithmetical mistakes. That is why motivation figures prominently in the literature on self-deception.

Elsewhere, I have distinguished between what I call *straight* and *twisted* cases of self-deception (Mele 1999, 2001). In straight cases, which have dominated the literature, people are self-deceived in believing something that they want to be true – for example, that their children are not using illegal drugs. In twisted cases, people are self-deceived in believing something that they want to be false (and do not also want to be true). For example, an insecure, jealous husband may believe that his wife is having an affair despite having only thin evidence of infidelity and despite his not wanting it to be the case that she is so engaged.

Some illustrations of ways in which our desiring that  $p$  can contribute to our believing that  $p$  in instances of straight self-deception will be useful (see Mele 2001, pp. 26-27). Often, two or more of the phenomena I describe are involved in an instance of self-deception.

1. *Negative Misinterpretation.* Our desiring that  $p$  may lead us to misinterpret as not counting (or not counting strongly) against  $p$  data that we would easily recognize to count (or count strongly) against  $p$  in the desire's absence. For example, Rex just received a rejection notice on a journal submission. He hopes that the rejection was unwarranted, and he reads through the referees' comments. Rex decides that the referees misunderstood two important but complex points and that their objections consequently do not justify the rejection. However, the referees' criticisms were correct, and a few days later, when Rex rereads his paper and the comments in a more impartial frame of mind, it is clear to him that this is so.

2. *Positive Misinterpretation.* Our desiring that  $p$  may lead us to interpret as *supporting*  $p$  data that we would easily recognize to count against  $p$  in the desire's absence. For example, Sid is very fond of Roz, a college classmate with whom he often studies. Because he wants it to be true that Roz loves him, he may interpret her declining his invitations to various social events and reminding him that she has a steady boyfriend as an effort on her part to "play hard to get" in order to encourage Sid to continue to pursue her and prove that his love for her approximates hers for him. As Sid interprets Roz's behavior, not only does it fail to count against the hypothesis that she loves him, it is evidence that she does love him. This contributes to his believing, falsely, that Roz loves him.

3. *Selective Focusing/Attending.* Our desiring that  $p$  may lead us to fail to focus attention on evidence that counts against  $p$  and to focus instead on evidence suggestive of  $p$ . Beth is a twelve-year-old whose father died recently. Owing partly to her desire that she was her father's favorite, she finds it comforting to attend to memories and photographs that place her in the spotlight of her father's affection and unpleasant to attend to memories and photographs that place a sibling in that spotlight. Accordingly, she focuses her attention on the former and is inattentive to the latter. This contributes to Beth's coming to believe – falsely – that she was her father's favorite child. In fact, Beth's father much preferred the company of her brothers, a fact that the family photo albums amply substantiate.

4. *Selective Evidence-Gathering.* Our desiring that  $p$  may lead us both to overlook easily obtainable evidence for  $\sim p$  and to find evidence for  $p$  that is much less accessible. For example, Betty, a political campaign staffer who thinks the world of her candidate, has heard rumors from the opposition that he is sexist, but she hopes he is not. That hope motivates her to scour his past voting record for evidence of his political correctness on gender issues and to consult people in her own campaign office about his personal behavior. Betty may miss some obvious, weighty evidence that her boss is sexist – which he in fact is – even though she succeeds in finding less obvious and less weighty evidence for her favored view. As a result, she may come to believe that her boss is not sexist. Selective evidence-gathering may be analyzed as a combination of hyper-sensitivity to evidence (and sources of evidence) for the desired state of affairs and blindness – of which there are, of course, degrees – to contrary evidence (and sources thereof).

In none of these examples does the person hold the true belief that  $\sim p$  and then intentionally bring it about that he or she believes that  $p$ . Yet, assuming that these people acquire relevant false, unwarranted beliefs in the ways described, these are garden-variety instances of self-deception.<sup>2</sup> Rex is self-deceived in believing that his article was wrongly rejected, Sid is self-deceived in believing certain things about Roz, and so on.

We can understand why, owing to her desire that her father loved her most, Beth finds it pleasant to attend to photographs and memories featuring her as the object of her father's affection and painful to attend to photographs and memories that put others in the place she prizes. But how do desires that  $p$  trigger and sustain the two kinds of misinterpretation and selective evidence-gathering? It is not as though these activities are intrinsically pleasant, as attending to pleasant memories, for example, is intrinsically pleasant.

Attention to some sources of *unmotivated* biased belief sheds light on this issue. Several such sources have been identified (see Mele 2001, pp. 28-31), including the following two:

---

<sup>2</sup> If, in the way I described, Betty acquires or retains the false belief that her boss is not sexist, it is natural to count her as self-deceived. This is so even if, owing to her motivationally biased evidence-gathering, the evidence that she actually *has* does not weigh more heavily in support of the proposition that her boss is sexist than against it.

1. *Vividness of information.* A datum's vividness for us often is a function of such things as its concreteness and its sensory, temporal, or spatial proximity. Vivid data are more likely to be recognized, attended to, and recalled than pallid data. Consequently, vivid data tend to have a disproportional influence on the formation and retention of beliefs.

2. *The confirmation bias.* People testing a hypothesis tend to search (in memory and the world) more often for confirming than for disconfirming instances and to recognize the former more readily (Baron 1988, pp. 259-65). This is true even when the hypothesis is only a tentative one (and not a belief one has). People also tend to interpret relatively neutral data as supporting a hypothesis they are testing (Trope et al. 1997, p. 115).

Although sources of biased belief apparently can function independently of motivation, they also may be triggered and sustained by desires in the production of *motivationally* biased beliefs.<sup>3</sup> For example, desires can enhance the vividness or salience of data. Data that count in favor of the truth of a proposition that one hopes is true may be rendered more vivid or salient by one's recognition that they so count. Similarly, desires can influence which hypotheses occur to one and affect the salience of available hypotheses, thereby setting the stage for the confirmation bias.<sup>4</sup> Owing to a desire that  $p$ , one may test the hypothesis that  $p$  is true rather than the contrary hypothesis. In these ways and others, a desire that  $p$  may help produce an unwarranted belief that  $p$ .

An interesting recent theory of lay hypothesis testing is designed, in part, to accommodate self-deception. I explore it in Mele 2001, where I offer grounds for caution and moderation and argue that a qualified version is plausible.<sup>5</sup> I call it the *FTL theory*, after the authors of the two articles on which I primarily drew, Friedrich 1993 and Trope and Liberman 1996. Here, I offer a thumbnail sketch.

The basic idea of the FTL theory is that a concern to minimize costly errors drives lay hypothesis testing. The *errors* on which the theory focuses are false beliefs. The *cost* of a false belief is the cost, including missed opportunities for gains, that it would be reasonable for the person to expect the belief – if false – to have, given his desires and beliefs, if he were to have expectations about such things. A central element of the FTL theory is a “confidence threshold” – or a “threshold,” for short. The lower the threshold, the thinner the evidence sufficient for reaching it. Two thresholds are relevant to each hypothesis: “The acceptance threshold is the minimum confidence in the truth of a hypothesis,”  $p$ , sufficient for acquiring a belief that  $p$  “rather than continuing to test [the hypothesis], and the rejection threshold is the minimum confidence in the untruth of a hypothesis,”  $p$ , sufficient for acquiring a belief that  $\sim p$  “and discontinuing

<sup>3</sup> I develop this idea in Mele 1987, ch. 10 and 2001. Kunda 1990 develops the same theme, concentrating on evidence that motivation sometimes primes the confirmation bias. Also see Kunda 1999, ch. 6.

<sup>4</sup> For motivational interpretations of the confirmation bias, see Friedrich 1993 and Trope and Liberman 1996, pp. 252-265.

<sup>5</sup> See Mele 2001, pp. 31-49, 63-70, 90-91, 96-98, 112-118.

the test” (Trope and Liberman 1996, p. 253). The two thresholds often are not equally demanding, and acceptance and rejection thresholds respectively depend “primarily” on “the cost of false acceptance relative to the cost of information” and “the cost of false rejection relative to the cost of information.” The “cost of information” is simply the “resources and effort” required for gathering and processing “hypothesis-relevant information” (Trope and Liberman 1996, p. 252).

Confidence thresholds are determined by the strength of aversions to specific costly errors together with information costs. Setting aside the latter, the stronger one’s aversion to falsely believing that  $p$ , the higher one’s threshold for belief that  $p$ . These aversions influence belief in a pair of related ways. First, because, other things being equal, lower thresholds are easier to reach than higher ones, belief that  $\sim p$  is a more likely outcome than belief that  $p$ , other things being equal, in a hypothesis tester who has a higher acceptance threshold for  $p$  than for  $\sim p$ . Second, the aversions influence *how* we test hypotheses – for example, whether we exhibit the confirmation bias – and *when we stop* testing them (owing to our having reached a relevant threshold).<sup>6</sup>

Friedrich claims that desires to avoid specific errors can trigger and sustain “automatic test strategies” (Friedrich 1993, p. 313), and this supposedly happens in roughly the nonintentional way in which a desire that  $p$  results in the enhanced vividness of evidence for  $p$ . In Mele 2001 (pp. 41- 49, 61-67), I argue that a person’s being more strongly averse to falsely believing that  $\sim p$  than to falsely believing that  $p$  may have the effect that he primarily seeks evidence for  $p$ , is more attentive to such evidence than to evidence for  $\sim p$ , and interprets relatively neutral data as supporting  $p$ , without this effect’s being mediated by a belief that such behavior is conducive to avoiding the former error. The stronger aversion may simply frame the topic in a way that triggers and sustains these manifestations of the confirmation bias without the assistance of a belief that behavior of this kind is a means of avoiding particular errors. Similarly, having a stronger aversion that runs in the opposite direction may result in a skeptical approach to hypothesis testing that in no way depends on a belief to the effect that an approach of this kind will increase the probability of avoiding the costlier error. Given the aversion, skeptical testing is predictable independently of the agent’s believing that a particular testing style will decrease the probability of making a certain error.

The FTL theory applies straightforwardly to both straight and twisted self-deception. Friedrich writes:

a prime candidate for primary error of concern is believing as true something that leads [one] to mistakenly criticize [oneself] or lower [one’s] self-esteem. Such costs are generally highly salient and are paid for immediately in terms

---

<sup>6</sup> Whether and to what extent subjects display the confirmation bias depends on such factors as whether they are given a neutral perspective on a hypothesis or, instead, the perspective of someone whose job it is to detect cheaters. See Gigerenzer and Hug 1992.

of psychological discomfort. When there are few costs associated with errors of self-deception (incorrectly preserving or enhancing one's self-image), mistakenly revising one's self-image downward or failing to boost it appropriately should be the focal error. (Friederich 1993, p. 314)

Here, he plainly has straight self-deception in mind, but he should not stop there. Whereas for many people it may be more important to avoid acquiring the false belief that their spouses are having affairs than to avoid acquiring the false belief that they are not so engaged, the converse may well be true of some insecure, jealous people. The belief that one's spouse is unfaithful tends to cause significant psychological discomfort. Even so, avoiding falsely believing that their spouses are faithful may be so important to some people that they test relevant hypotheses in ways that, other things being equal, are less likely to lead to a false belief in their spouses' fidelity than to a false belief in their spouses' infidelity. Furthermore, data suggestive of infidelity may be especially salient for these people and contrary data quite pallid by comparison. Don Sharpsteen and Lee Kirkpatrick observe that "the jealousy complex" – that is, "the thoughts, feelings, and behavior typically associated with jealousy episodes" – is interpretable as a mechanism "for maintaining close relationships" and appears to be "triggered by separation, or the threat of separation, from attachment figures" (Sharpsteen and Kirkpatrick 1997, p. 627). It certainly is conceivable that, given a certain psychological profile, a strong desire to maintain one's relationship with one's spouse plays a role in rendering the potential error of falsely believing one's spouse to be innocent of infidelity a "costly" error, in the FTL sense, and more costly than the error of falsely believing one's spouse to be guilty. After all, the former error may reduce the probability that one takes steps to protect the relationship against an intruder. The FTL theory provides a basis for an account of both straight and twisted self-deception (Mele 2001, ch. 5).

I have concentrated on roles for motivation in self-deception. A brief comment on roles for affect is in order (see Mele 2003). It is often held that emotions have desires as constituents. Even if that is so, might emotions contribute to some instances of self-deception in ways that do not involve a constituent desire's making a contribution? Suppose that Art is angry at Bob for a recent slight. His anger may prime the confirmation bias by suggesting an emotion-congruent hypothesis about Bob's current behavior – for example, that Bob is behaving badly again – and it may increase the salience of data that seem to support that hypothesis.<sup>7</sup> There is evidence that anger tends to focus attention selectively on explanations in terms of "agency," as opposed to situational factors (Keltner et al. 1993). Perhaps Art's anger leads him to view Bob's behavior as more purposeful and more indicative of a hostile intention than he otherwise would. If anger has a desire as a constituent, it is, roughly, a desire to lash out against the target of one's

<sup>7</sup> There is evidence that "emotional states facilitate the processing of congruent stimuli" and that "attentional processes are involved in [this] effect" (Derryberry 1988, pp. 36, 38). Gordon Bower and Joseph Forgas review evidence that emotions make "emotionally congruent interpretations of ambiguous stimuli more available" (2000, p. 106).

anger. Possibly, anger can play the biasing roles just mentioned without any constituent desire's playing them.

If an emotion can do this, perhaps an emotion may contribute to an instance of self-deception that involves *no* desires at all as significant biasing causes. It is conceivable, perhaps, that Art enters self-deception in acquiring the belief that Bob is behaving badly now, that the process that results in this belief features his anger's playing the biasing roles just described, and that no desires of his have a biasing effect in this case. If it is assumed that Art believes that Bob is behaving badly despite having stronger evidence for the falsity of that hypothesis than for its truth, an FTL theorist will find it plausible that Art had a lower threshold for acceptance of that hypothesis than for rejection of it, that the difference in thresholds is explained at least partly in terms of relevant desires, and that this difference helps to explain Art's acquiring the belief he does. But this position on Art's case is debatable, and I leave the matter open.

Although I have never offered a conceptual analysis of self-deception, I have suggested the following proto-analysis: people enter self-deception in acquiring a belief that *p* if and only if *p* is false and they acquire the belief in a *suitably biased way* (Mele 2001, p. 120).<sup>8</sup> The suitability at issue is a matter of kind of bias, degree of bias, and the nondeviance of causal connections between biasing processes (or events) and the acquisition of the belief that *p*.<sup>9</sup> I suggest that, as self-deception is commonly conceived, something along the following lines is a test for a level of motivational or emotional bias appropriate to a person's being self-deceived in acquiring a belief that *p*: given that *S* acquires a belief that *p* and *D* is the collection of relevant data readily available to *S* during the process of belief-acquisition, if *D* were made readily available to *S*'s impartial cognitive peers and they were to engage in at least as much reflection on the issue as *S* does and at least a moderate amount of reflection, those who conclude that *p* is false would significantly outnumber those who conclude that *p* is true. Call this *the impartial observer test*.<sup>10</sup> It is a test for a person's satisfying the suitable bias condition on self-deception. A person's passing the test is evidence of bias suitable for self-deception.

I close this section with some comments on this test. By "cognitive peers," I mean people who are very similar to the person being tested in such things as education and intelligence. Cognitive peers who share certain relevant desires with the subject – as one's spouse may share one's desire that one's child is not using illegal drugs – may

---

<sup>8</sup> The requirement that *p* be false is purely semantic. By definition, one is *deceived in* believing that *p* only if *p* is *false*; the same is true of being *self-deceived in* believing that *p*. The requirement does not imply that *p*'s being false has special importance for the *dynamics* of self-deception. Biased treatment of data may sometimes result in someone's believing an improbable proposition, *p*, that happens to be *true*. There may be self-deception in such a case, but the person is not self-deceived in believing that *p*, nor in acquiring the belief that *p*. On a relevant difference between being *deceived in* believing that *p* and being *deceived into* believing that *p*, see Mele 1987, pp. 127-28.

<sup>9</sup> On deviant and nondeviant causation in this connection, see Mele 2001, pp. 121-23.

<sup>10</sup> This is a modified version of the test suggested in Mele 2003, p. 164. Discussion with Charles Hermes and Brian McLaughlin motivated the modifications.

often acquire the same unwarranted belief that the subject does, given the same data. But the relevant cognitive peers, for present purposes, are *impartial* observers. At least a minimal requirement for impartiality in the present context is that one neither share the subject's desire that  $p$  nor have a desire that  $\sim p$ . Another plausible requirement is that one not prefer avoidance of either of the following errors over the other: falsely believing that  $p$  and falsely believing that  $\sim p$ . A third is that one not have an emotional stake in  $p$ 's truth or falsity. The test is a test for a level of motivational or emotional bias appropriate to self-deception. I take the suitability of the impartial observer test – or something similar, at least – to be implicit in the conceptual framework that informs common-sense judgments about what is and is not plausibly counted as an instance of self-deception.<sup>11</sup>

## 2. Delusion

In this section, I treat the terms “self-deception” and “delusion” as labels for phenomena and ask how the phenomena are related. Of course, identity is a relation, as is mutual exclusion. My question is not meant to be a loaded one.

DSM-IV offers the following gloss on “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. The belief is not one ordinarily accepted by other members of the person's culture or subculture (e.g., it is not an article of religious faith). When a false belief involves a value judgment, it is regarded as a delusion only when the judgment is so extreme as to defy credibility. Delusional conviction occurs on a continuum and can sometimes be inferred from an individual's behavior. It is often difficult to distinguish between a delusion and an overvalued idea (in which case the individual has an unreasonable belief or idea but does not hold it as firmly as is the case with a delusion). (1994, p. 765)

Two points merit emphasis. As delusions are understood in DSM-IV, they are exceptionally resistant to contrary evidence and the contrary evidence is very strong. Both points are reinforced elsewhere in DSM-IV: “The distinction between a delusion and a strongly held idea . . . depends on the degree of conviction with which the belief is held despite clear contradictory evidence” (p. 275). I take it that “degree of conviction” (or firmness of belief) here is at least partly a matter of how strong the contrary evidence would need to be to undermine the belief.

<sup>11</sup> I say “or something similar” because, for reasons that emerge in section 2, the test, as formulated, may not be reliable in unusual cases of certain kinds.

The idea that all delusions are “based on incorrect inference about external reality” is dispensable. A person might have the delusion that he lacks certain internal organs (Davies et al. 2001, p. 136). Presumably, such a delusion need not be based on an inference about external reality. With this exception, I follow the quoted gloss.

One way to approach the connection between self-deception and delusion features intuitions about cases. Another approach features an investigation of the causes, in these spheres, of the pertinent beliefs. I opt for the latter. I will consider three delusions: the Capgras delusion, delusional jealousy (or the Othello syndrome), and the reverse Othello syndrome.

## 2.1. The Capgras Delusion

Carl believes that his wife has been “replaced by an exact replica or impostor” (Stone and Young 1997, p. 327). This is an instance of the Capgras delusion. Part of the cause in Carl’s case, apparently, is a brain injury that deprives him of his normal affective response to his wife’s face (Stone and Young 1997, p. 337). Various views have been advanced about additional causal contributors.

Brendan Maher’s model of delusions includes the following two hypotheses:

1. Delusional beliefs, like normal beliefs, arise from an attempt to explain experience.
2. The processes by which deluded persons reason from experience to belief are not significantly different from the processes by which non-deluded persons do. (Maher 1999, p. 550)

Carl has a new way of experiencing his wife’s face. On Maher’s view, Carl’s delusional belief is a product of his attempt to explain this, and his reasoning is not significantly different from normal reasoning. The claim about Carl’s reasoning suggests that most normal people who are presented with the evidence Carl has would come to his conclusion about it. Is it true that if Carl’s new way of experiencing his wife’s face were made clear to non-deluded people, most would infer that she has been replaced by an impostor? Certainly, Carl’s doctors do not infer this, and it seems that normal people would find the impostor hypothesis about as far-fetched as his doctors do. Also, there is evidence that some people with “the same kind of experience of faces as Capgras patients” do not have the Capgras delusion (Davies et al. 2001, p. 144).

Considerations such as these have been used to motivate a two-factor alternative to Maher’s model of delusion. Davies et al. suggest two possibilities for a second factor and express a preference for the following idea. Carl experiences his wife as someone who looks just like her but is not really her, and he “accepts this perceptual experience as veridical” (Davies et al. 2001, p. 153). The first factor is this experience, which includes the impostor idea as part of its content, as opposed, for example, to the idea’s being a

hypothesis that is separate from and prompted by the experience. The second factor is a problem that accounts for Carl's accepting the experience as veridical rather than rejecting it as not veridical. The main proposal Davies et al. offer about the form this problem takes is intriguing. As they observe, "Normal subjects are . . . able to suspend their unreflective acceptance of veridicality and make a more detached and critical assessment of the credentials of their perceptual experiences" (Davies et al. 2001, p. 153). Their proposal is that Capgras patients have a deficit in this connection. If this is "the nature of the second factor in the etiology of delusions," then hypotheses that are included in "the patients' own perceptual experience [are] resistant to being critically assessed and recognized as implausible, but hypotheses generated by someone else [are] assessed in the normal way" (Davies et al. p. 153).

Davies et al. recognize that their proposal generates the prediction that people with this deficit will be led to have false beliefs by their visual illusions in general, and they are clearly uncomfortable about this (Davies et al. 2001, p. 153). My own immediate concern is with the bearing of their proposal and Maher's proposal on the impartial observer test. I start with the latter proposal. As I mentioned, it certainly seems to suggest that most normal people who are presented with the evidence the Capgras patient has would come to the patient's conclusion about it – that a loved one has been replaced by an impostor. This suggestion seems so implausible that one wonders exactly what Maher meant. Suppose he believes that it cannot actually be made clear to normal people what the Capgras patient's pertinent experience is like. Then he may say that people who lack that experience cannot actually have the pertinent evidence. This would limit members of the panel, for the purposes of the impartial observer test, to people who have experiences of the sort characteristic of Capgras patients. Exclude all such people who do not satisfy the conditions for membership on the panel. What would the majority of the remainder conclude?

As Davies et al. observe, "At least some delusional patients show considerable appreciation of the implausibility of their delusional beliefs" (Davies et al. 2001, p. 149). Andrew Young writes: "Capgras delusion patients can be . . . able to appreciate that they are making an extraordinary claim. If you ask 'what would you think if I told you my wife had been replaced by an impostor,' you will often get answers to the effect that it would be unbelievable, absurd, an indication that you had gone mad" (Young 1998, p. 37). Even many delusional patients on the panel might judge that Carl's wife was not replaced by an impostor. And a higher percentage of panel members with similar experiences but no delusions might make that judgment. Suppose the overwhelming majority of panelists deem Carl's belief false. Would that constitute good evidence that Carl's treatment of data is motivationally or emotionally biased?

The basic question behind the impartial observer test, of course, is whether something in the motivation/emotion category biased the subject's treatment of data in the process that produced the belief at issue and whether, if this happened, the biasing was robust

enough to be appropriate for self-deception. The idea is to strip away the potential motivational and emotional sources of bias while holding the evidence fixed and to see what happens. If the subject's belief is reached by the great majority in the absence of those sources, that is evidence that they did not play the biasing role at issue in the subject. If the converse belief is reached by the great majority, that is evidence that motivation or emotion did play this biasing role in the subject, and the relative size of the majority is evidence of the robustness of the role. But, of course, some nonmotivational and nonemotional factor might be present in the subject in the latter case and absent in the panel, and it might be doing a great deal of causal or explanatory work. This is exactly the situation with Capgras patients if what Davies et al. propose is correct. That is, what would account for the difference in belief is a certain *cognitive* deficit that is outside the categories of motivation and emotion. And even if it were insisted that people must have that deficit in order to count as *cognitive* peers of the target person, that would make no difference; for the proposed deficit shows up only in responses to one's own experiences.<sup>12</sup>

If what produces the Capgras delusion is a weird experience together with the removal or disabling of a cognitive mechanism that, in special cases, inhibits a kind of default transition from experience to corresponding belief, the delusion seems to lie well beyond the sphere of self-deception. And independently of the proposal by Davies et al. if we lack good reason to believe that motivation or emotion biases the Capgras patient's treatment of data, thereby contributing to the delusional belief, we lack good reason to believe that the delusion is an instance of self-deception. Notice that accepting that the Capgras delusion is explained partly by emotional factors does not commit one to accepting that emotion biases the person's treatment of data. For example, we apparently should accept that a major emotional change – a certain loss of affect – plays an important role in producing the delusion. But this loss is a *cause* of relevant experiential data: *causing* data is one thing; *biasing* a person's treatment of data is another.

Young reports on a “person who experienced both the Cotard and Capgras delusions in sequence” (Young 1999, p. 577). People with the former delusion believe that they themselves are dead. Young writes:

This curious association of two unusual delusions has been reported in other cases too, and the key factor seems to be the patients' moods – when in a suspicious mood, they think that other people are impostors, when depressed they think they are dead. There is an obvious parallel here to... findings that people with persecutory delusions tend to make external attributions and depressed people internal attributions as to the causes of negative events. (Young 1999, p. 577)

---

<sup>12</sup> Recall my assertion in section 1 that “a person's passing the [impartial observer] test is evidence of bias suitable for self-deception?” One moral of the paragraph to which this note is appended is that if a special cognitive deficit of the kind at issue is doing the causal or explanatory work, that fact undermines the evidence.

What might Davies et al. say about this? Perhaps, that just as the Capgras patient's experience includes the impostor idea as part of its content, the Cotard patient's experience includes the idea that the subject is dead as part of its content. Perhaps in people with both delusions at different times, their feelings of suspicion are part of the cause of their having an experience that includes the impostor content, and their depression is part of the cause of their having an experience that includes the "I am dead" content. If so, affective states – depression and feelings of suspicion – would help to explain the delusions. But again they would do so by helping to cause experiential data – these experiences with strange content – rather than by biasing the person's treatment of data. My question is whether the Capgras patient's treatment of relevant data is motivationally or emotionally biased. The evidence and theorizing that I have seen does not support an affirmative answer.

## 2.2. Delusional jealousy

Next on the agenda is delusional jealousy, one of the types of delusion identified in DSM-IV's gloss on delusion. It is defined there as "the delusion that one's sexual partner is unfaithful" (1994, p. 765). David Enoch asserts that it is difficult to differentiate "between normal and excessive, excessive and neurotic, and neurotic and psychotic [jealousies]. The various types overlap and the boundaries are blurred" (Enoch 1991, p. 52). In section 1, I offered a scenario featuring a jealous husband as an illustration of twisted self-deception. Enoch's assertion suggests that, in the sphere of jealousy, one might be able to locate self-deception on a continuum that includes delusional jealousy, and that being self-deceived in believing that one's sexual partner is unfaithful might at least overlap with delusional jealousy. This suggestion is consistent with DSM-IV's description of the "jealous type" of delusional disorder: "This subtype applies when the central theme of the person's delusion is that his or her spouse or lover is unfaithful. This belief is arrived at without due cause and is based on incorrect inference supported by small bits of 'evidence' (e.g., disarrayed clothing or spots on the sheets), which are collected and used to justify the delusion" (1994, p. 297).

There are also grounds for pessimism about the suggestion at issue. Michael Soyka observes that "delusions of jealousy are a frequent symptom in various psychiatric disorders. ... Most ... patients with delusions of infidelity are schizophrenics..." (Soyka 1995, p. 118). Barbara Breitner and David Anderson report that "three large studies found 30–50% of the morbidly jealous suffered from psychosis, a similar proportion neurosis or personality disorder, 5–7% alcoholism and the remainder miscellaneous conditions, most commonly organic disorders" (Breitner and Anderson 1994, p. 703). Silva et al. assert that "delusional jealousy rarely exists as the only prominent symptom but is usually found in conjunction with other symptoms, including other delusions and psychotic symptoms" (Silva et al. 1998, p. 616). In a study of 20 people with delusional jealousy, half had directly relevant auditory hallucinations (some of which were commands to attack the partner)

and two had relevant visual hallucinations (Silva et al. 1998, pp. 615-16). In a study of 133 demented patients, “all patients with delusional jealousy ... had at least one other psychotic symptom,” as compared with “70.5% of patients without delusional jealousy” (Tsai et al. 1997, p. 492).

One possibility is that although jealous people who are self-deceived in believing that their partners are unfaithful and people with delusional jealousy believe the same thing, the causes of that belief in the two groups are so different that the groups do not overlap. Consider people with delusional jealousy who have auditory hallucinations informing them that their partners are unfaithful or visual hallucinations of their partners being unfaithful. A proposal like the one Davies et al. make about Capgras patients may be made about them. Perhaps, owing to a cognitive deficit, they accept the “experience as veridical.” In those without such hallucinations, one needs to look elsewhere for causes. Tsai et al. found in their study of 133 demented patients that “the frequency of delusions of theft [and] persecutory delusions ... was significantly higher in the delusional jealousy group” (Tsai et al. 1997, p. 492).<sup>13</sup> When delusions show up in pairs or larger groups, one is inclined to look for a common cause, especially when the delusions are thematically related. Infidelity may be viewed as encompassing both theft (by the new romantic partner or partners) and persecution. To the extent to which one is inclined to see delusions of theft and persecution as falling outside the sphere of self-deception and as being explained in part by a cognitive deficit, one should have the same inclination toward delusions of infidelity in people who have one or both of the other delusions.

Another possibility merits attention – namely, that the FTL model applies straightforwardly to some people with delusional jealousy, perhaps especially those in whom this *is* “the only prominent symptom”. The hypothesis is that the error costs for some people with this problem are such that they have an extremely high threshold for accepting the fidelity proposition. It is extremely important to them not to believe that their partners are faithful if, in fact, they are unfaithful. Adequate support for this hypothesis would require a plausible account of why it is that they have these extreme error costs. Such an account would provide grounds for belief that motivation is playing a major biasing role and that these delusionally jealous people are self-deceived.

### 2.3. Reverse Othello Syndrome

Reverse Othello syndrome is “delusional belief in the fidelity of a romantic partner” (Butler 2000, p. 85). As in ordinary straight self-deception, the person believes something that he wants to be true. Indeed, a stock example of straight self-deception is the person who believes in the fidelity of his or her spouse despite strong evidence to the contrary – evidence that would lead the great majority of impartial cognitive peers to

---

<sup>13</sup> They also mention visual hallucinations and Capgras syndrome in this sentence.

believe in infidelity. Accordingly, the prospects for an important biasing role for motivation in this syndrome might look bright.

Peter Butler examines the case of a middle-aged man, B. X., who suffered a severe head injury in a high-speed car accident. His romantic partner, N, ended their relationship five months later, which B. X. acknowledged. But, despite the absence of contact with her, he subsequently “developed an intense delusional belief that [she] remained sexually faithful and continued as his lover and life partner” (Butler 2000, p. 86). He even came to believe that he married N while he was a patient (Butler 2000, p. 87). Doctors tested B. X. for other delusions and found no unrelated ones (Butler 2000, p. 88). After some months, “his delusional system began to break up.” A few months later he accepted the truth.

One important difference between B. X. and his self-deceived counterpart in the stock example I mentioned is B. X.’s belief that he married N. If there is any experiential basis for his belief in the marriage, it is something on the order of dreams or hallucinations. B. X. reported that the wedding “occurred at the Central Synagogue in front of several hundred guests” (Butler 2000, p. 88). He might have dreamed or hallucinated that. Suppose he did. And suppose the dream or hallucination – possibly a repeated one – was a cause of his belief and was caused in part by a wish to be married to N or some wish of that kind. Then motivation played a role in B. X.’s belief in his marriage. But its playing this particular role would highlight a role for a serious cognitive deficit. When a dream or hallucination is radically out of line with obvious reality, people without a serious cognitive deficit do not regard the experience as veridical after they awake or exit the hallucination.

Butler reports that “when questioned about the absence of photographs of the ceremony or corroboration from his family [B. X.] remained adamant the marriage had occurred and set his communicator to repeat the words ‘just because’” (Butler 2000, p. 88). Seemingly, B. X. wants not to think about these absences. He may understandably be motivated to focus his attention on the imagined marriage and to ignore considerations that point to its being a fantasy. The belief that he is married to N obviously gives B. X. pleasure, and entertaining challenges to that belief is unpleasant for him. Selective focusing or attending, which is at work in some ordinary cases of self-deception, may also be at work in B. X. Even if he does not enter self-deception in acquiring the belief that he is married to N, he may be self-deceived in continuing to believe this.

What about people with the Capgras delusion? Might they be self-deceived in persisting in believing that a loved one has been replaced by an impostor? Recall the assertion by Davies et al. that “normal subjects are ... able to suspend their unreflective acceptance of veridicality and make a more detached and critical assessment of the credentials of their perceptual experiences” (Davies et al. 2001, p. 153). Suppose that people with the Capgras delusion are literally *unable* to do this. Then even if, like B. X., they refuse to reflect on challenges to their beliefs raised in conversation, this is not a

cause of the persistence of their delusional beliefs. For even if they were to reflect on the challenges, no change of belief would result; they are, by hypothesis, unable to shed the beliefs. Whether these people are self-deceived in retaining their delusional beliefs depends on the causes of their retention of them. If selective focusing is present here but is not a cause of belief retention, the observation that it is present does not warrant a diagnosis of self-deception.

Suppose that the pertinent cognitive deficit in some Capgras patients does not render them unable “to suspend their unreflective acceptance of veridicality” and instead makes it extremely difficult for them to do this. Then processes like selective focusing might do some work in sustaining the delusional belief. But the causal contribution may be so small that we may be disinclined to count the Capgras patient as self-deceived.

Should I apologize for my inconclusiveness about the delusions I have discussed? I have come to believe that owing to the difficulty of the subject matter, an apology is not in order. But perhaps I have deceived myself.<sup>14</sup>

#### REFERENCES

- American Psychiatric Association. (1994), *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.), Washington, D.C.
- Baron J. (1988), *Thinking and Deciding*, Cambridge: Cambridge University Press
- Bower G. & J. Forgas (2000), “Affect, Memory, and Social Cognition”, in Eich E., Kihlstrom J., Bower G., Forgas J. & P. Niedenthal, 2000, pp. 87-168
- Breitner B. & D. Anderson (1994), “The Organic and Psychological Antecedents of Delusional Jealousy in Old Age”, *International Journal of Geriatric Psychiatry*, 9, pp. 703-707
- Butler P. (2000), “Reverse Othello Syndrome Subsequent to Traumatic Brain Injury”, *Psychiatry*, 63, pp. 85-92
- Charlton B. & H. McClelland (1999), “Theory of Mind and Delusional Disorders”, *The Journal of Nervous and Mental Disease*, 187, pp. 380-383
- Davies M., Langdon R. & N. Breen (2001), “Monothematic Delusions: Towards a Two-Factor Account”, *Philosophy, Psychiatry, and Psychology*, 8, pp. 133-158
- Derryberry D. (1988), “Emotional Influences on Evaluative Judgments: Roles of Arousal, Attention, and Spreading Activation”, *Motivation and Emotion*, 12, pp. 23-55
- Eich E., Kihlstrom J., Bower G., Forgas J. & P. Niedenthal eds. (2000), *Cognition and Emotion*, Oxford: Oxford University Press
- Enoch D. (1991), “Delusional Jealousy and Awareness of Reality”, *British Journal of Psychiatry*, 159, (suppl. 14), pp. 52-56

---

<sup>14</sup> A draft of this article was presented at a conference at the Macquarie Centre for Cognitive Science in November, 2004. I am grateful to my audience there for useful feedback. This article will appear in the conference proceedings: Fernandez J. and T. Bayne, eds., *Delusions, Self-Deception, and Affective Influences on Belief-Formation* (Psychology Press, forthcoming). The article is published here with the kind permission of Psychology Press.

- Friedrich J. (1993), "Primary Error Detection and Minimization PEDMIN Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena", *Psychological Review*, 100, pp. 298-319
- Gigerenzer G. & K. Hug (1992), "Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change", *Cognition*, 43, pp. 127-171
- Hatzimoysis A. ed. (2003) *Philosophy and the Emotions*, Cambridge: Cambridge University Press
- Higgins E.T. & A. Kruglanski eds. (1996), *Social Psychology: Handbook of Basic Principles*, New York: Guilford Press
- Keltner D., Ellsworth P. & K. Edwards (1993), "Beyond Simple Pessimism: Effects of Sadness and Anger on Social Perception", *Journal of Personality and Social Psychology*, 64, pp. 740-752
- Kunda Z. (1990), "The Case for Motivated Reasoning", *Psychological Bulletin*, 108, pp. 480-498
- Kunda Z. (1999), *Social Cognition*, Cambridge, Mass.: MIT Press
- Maher B. A. (1999), "Anomalous Experience in Everyday Life: Its Significance for Psychopathology", *The Monist*, 82, pp. 547-570
- Mele A. (1987), *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*, New York: Oxford University Press
- Mele A. (1999), "Twisted Self-Deception", *Philosophical Psychology*, 12, pp. 117-137
- Mele A. (2001), *Self-Deception Unmasked*, Princeton: Princeton University Press
- Mele A. (2003), "Emotion and Desire in Self-Deception", in Hatzimoysis 2003, pp. 163-179
- Myslobodsky M. ed. (1997), *The Mythomanias: The Nature of Deception and Self-Deception*, Mahwah, NJ: Lawrence Erlbaum
- Sharpsteen D. & L. Kirkpatrick (1997), "Romantic Jealousy and Adult Romantic Attachment", *Journal of Personality and Social Psychology*, 72, pp. 627-640
- Silvia J., Ferrari M., Leong G. & G. Penny (1998), "The Dangerousness of Persons with Delusional Jealousy", *Journal of the American Academy of Psychiatry and the Law*, 26, pp. 607-623
- Soyka M. (1995), "Prevalence of Delusional Jealousy in Schizophrenia", *Psychopathology*, 28, pp. 118-120
- Soyka M., Naber G. & A. Völcker (1991), "Prevalence of Delusional Jealousy in Different Psychiatric Disorders: An Analysis of 93 Cases", *British Journal of Psychiatry*, 158, pp. 549-553
- Stone T. & A. Young (1997), "Delusions and Brain Injury: The Philosophy and Psychology of Belief", *Mind and Language*, 12, pp. 327-364
- Trope Y. & A. Liberman (1996), "Social Hypothesis Testing: Cognitive and Motivational Mechanisms", in Higgins & Kruglanski 1996, pp. 239-270
- Trope Y., Gervy B. & N. Liberman (1997), "Wishful Thinking from a Pragmatic Hypothesis-Testing Perspective", in Myslobodsky 1997, pp. 105-131
- Tsai S.J., Hwang J.P., Yang C.H. & K.M. Liu (1997), "Delusional Jealousy in Dementia", *Journal of Clinical Psychiatry*, 58, pp. 492-494
- Young A. (1998), *Face and Mind*, Oxford: Oxford University Press
- Young A. (1999), "Delusions", *The Monist* 82, pp. 571-589

---

Received: September 22, 2006

Accepted: November 15, 2006

Florida State University  
 Dodd Hall  
 Tallahassee, FL 32306-1500  
 USA  
 almele@mailier.fsu.edu