

## EMOTION AND DESIRE IN SELF-DECEPTION

Alfred R. Mele, Florida State University

According to a traditional view of self-deception, the phenomenon 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.

Elsewhere (most recently in Mele 2001), I have criticized the traditional conception of self-deception and defended 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 contradictory propositions. Indeed, I have argued that garden-variety instances of self-deception do not include any of these things. On my view, to put it simply, 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.<sup>2</sup> Obviously, this shoulders me with the burden of showing what suitable bias amounts to, and I have had a lot to say about that. 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>3</sup>

In Mele 2001 (pp. 106-12), I suggested a test for relevant bias. I called it “the impartial observer test,” and I argued that its appropriateness is underwritten by the ordinary concept of self-deception. Here is an improved version: If S is self-deceived in believing that 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 (including merely hypothetical people) 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.<sup>4</sup> This is a test for the satisfaction of a necessary condition of being self-deceived in believing that p. One requirement for impartiality in the present context is that one neither desire that p nor desire that  $\sim p$ . Another is that one not prefer avoidance of either of the following errors over the other: falsely believing that p and falsely believing that  $\sim p$ . The kind of bias at issue may broadly be termed “motivational or emotional bias.” Although I have discussed biasing causes and processes – especially motivational ones – at length, I have left it open that a motivationally biased treatment of data is not required for self-deception and that emotions sometimes do the biasing work without motivation’s playing a biasing role. This is one of the two possibilities that I explore in this essay. The other is a more moderate thesis about the place of emotion in self-deception.

### 1. Background: Biased Belief and Self-Deception

In the present section, after briefly describing some mechanisms relevant to the production of motivationally biased belief of a sort appropriate to self-deception, I sketch a general account of such belief. My primary aim is to prepare the way for my discussion of emotionally biased belief in Section 2. I have reviewed empirical evidence of motivationally biased belief elsewhere (most recently in Mele 2001) and I will not do so again here.

Attention to some phenomena that have been argued to be sources of unmotivated or “cold” biased belief sheds light on motivationally biased belief. A number of such sources have been identified in the psychological literature, including the following two.

1. Vividness of information. A datum’s vividness for a person often is a function of the person’s interests, the concreteness of the datum, its “imagery-provoking” power, or its sensory, temporal, or spatial proximity (Nisbett and Ross 1980, p. 45). 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; Klayman and Ha 1987; Nisbett and Ross, pp. 181-82). This is true even when the hypothesis is only a tentative one (as opposed, e.g., to 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 may also be triggered and sustained by desires in the production of particular motivationally biased beliefs.<sup>5</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>6</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 contribute to the acquisition of an unwarranted belief that *p*.

Sometimes we generate our own hypotheses, and sometimes others suggest hypotheses to us – including extremely unpleasant ones. If we were consistently to concentrate primarily on confirmatory instances of hypotheses we are testing, independently of what is at stake, that would indicate the presence of a cognitive tendency or disposition that uniformly operates independently of desires. For example, it would indicate that desires never play a role in influencing the proportion of attention we give to evidence for the falsity of a hypothesis. However, there is powerful evidence that the “confirmation bias” is much less rigid than this. For example, in one study (Gigerenzer and Hug 1992), two groups of subjects are asked to test “social-contract rules such as ‘If someone stays overnight in the cabin, then that person must bring along a bundle of firewood . . .’” (Friedrich 1993, p. 313). The group asked to adopt “the perspective of a cabin guard monitoring compliance” showed an “extremely high frequency” of testing for disconfirming instances. The other group, asked to “take the perspective of a visitor trying to determine” whether firewood was supplied by visitors or a local club, displayed the common confirmation bias.

An interesting recent theory of lay hypothesis testing is designed, in part, to account for motivationally biased belief. I examined it in Mele 2001, where I offered grounds for caution and moderation and argued that a qualified version is plausible.<sup>7</sup> I named it the “FTL theory,” after the authors of the essays on which I primarily drew, Friedrich 1993 and Trope and Liberman 1996. I will offer a thumbnail sketch of the theory shortly. First, an explicit application of it to self-deception should be noted.

On James Friedrich's PEDMIN – “primary error detection and minimization” – model of lay hypothesis testing, “detection and minimization of crucial errors is . . . the central organizing principle” in this sphere (1993, p. 299). Regarding 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 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. (p. 314)

The basic idea of the FTL theory is that lay hypothesis testing is driven by a concern to minimize making costly errors. The errors in question 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 the notion of 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 producing a belief that  $p$ ; and “the rejection threshold is the minimum confidence in the untruth of a hypothesis,”  $p$ , sufficient for producing a belief that  $\sim p$  (Trope and Liberman 1996, p. 253). Acquiring the belief terminates hypothesis testing. The two thresholds often are not equally high, and the 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” (p. 252).

Confidence thresholds are determined by the strength of desires to avoid specific costly errors together with information costs. Setting aside the latter costs, the stronger one's desire to avoid falsely believing that  $p$ , the higher one's threshold for belief that  $p$ . These desires influence belief in two 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 desires at issue influence how we test hypotheses, not just when we stop testing them (owing to our having reached a relevant threshold). Recall the study in which subjects asked to adopt “the perspective of a cabin guard” showed an “extremely high frequency” of testing for disconfirming instances whereas subjects asked to “take the perspective of a visitor” showed the common confirmation bias.

It might be claimed that if avoidance desires of the kind under discussion function in the second way, they function in conjunction with beliefs to the effect that testing-behavior of a specific kind will tend to help one avoid making the costly errors at issue. It might be claimed, accordingly, that the pertinent testing-behavior is performed for a reason constituted by the desire and an instrumental belief of the kind just mentioned and that this behavior is therefore performed with the intention of trying to avoid, the pertinent error. The thrust of these claims is that the FTL theory accommodates the confirmation bias, for example, by invoking a model of intentional action.

This is not a feature of the FTL model, as its proponents understand it. Friedrich claims that desires to avoid specific errors can trigger and sustain “automatic test strategies” (p. 313), which supposedly happens in roughly the nonintentional way in which a desire that  $p$  enhances the vividness of evidence for  $p$ . A person’s having a stronger desire to avoid falsely believing that  $\sim p$  than to avoid 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 desire 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 a certain error. Similarly, having a stronger desire 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 stronger desire, skeptical testing is predictable independently of the agent’s believing that a particular testing style will decrease the probability of making a certain error. So at least I have argued elsewhere (Mele 2001, pp. 41-49, 61-67).

I will not defend this thesis again here. Nor am I claiming that the FTL theory is acceptable without qualification. The theory may accurately describe what happens in some or many cases of lay hypothesis testing that results in belief, and in many or all cases of self-deception.

One more piece of background is in order. Elsewhere, I have distinguished between “straight” and “twisted” self-deception (Mele 1997b; 1999; 2000; 2001, pp. 4-5, 94-118). In straight instances, we are self-deceived in believing something that we want to be true. In twisted instances, we are self-deceived in believing something that we want to be false (and do not also want to be true). Twisted self-deception may be exemplified by an insecure, jealous husband who believes that his wife is having an affair despite possessing only relatively weak evidence for that proposition and unambivalently wanting it to be false that she is so engaged.<sup>8</sup>

The FTL theory applies straightforwardly to twisted self-deception. 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 the relevant hypothesis 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. Don Sharpsteen and Lee Kirkpatrick observe that “the jealousy complex” – that is, “the thoughts, feelings, and behavior typically associated with jealousy episodes” – can be regarded as a mechanism “for maintaining close relationships” and appears to be “triggered by separation, or the threat of separation, from attachment figures” (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 a plausible account of twisted self-deception (see Mele 1999 and 2001, ch. 5).

## 2. Emotions in Self-Deception

I turn to possible roles for emotions in self-deception. Insofar as emotions are causes of belief-biasing desires or are partially constituted by such desires, they have a clear bearing on self-deception, if the FTL theory is on the right track. Consider Bob, who is self-deceived in believing that his wife, Ann, is not having an affair. Bob's love for Ann, or his fear that he cannot get along without her, may be a partial cause of his desire that she is not having an affair and, thereby, of his being self-deceived about this. If that desire increases the salience of his apparent evidence of her fidelity or helps shape his relevant confidence thresholds, emotions that contribute to the desire play an indirect part in this. Furthermore, Bob may fear that Ann is guilty of infidelity, and if a constituent of his fear is a desire that she is innocent, then the role the desire plays in his self-deception may be attributed to the fear, insofar as the fear encompasses the desire.<sup>9</sup>

There are additional possibilities. Obviously, people are averse to anxious feelings. Such feelings may be caused by reflection. In some cases, a desire that one's anxiety subside may play a role in attenuating or halting reflection on an unpleasant hypothesis, thereby decreasing the likelihood of one's undermining a contrary hypothesis to which one is attracted. Anxiety is an emotion: here, again, an emotion has an indirect role in potential self-deception. Some emotions may also help to explain some instances of self-deception by weakening one's motivation to assess evidence carefully (see Forgas 1995, p. 50), thereby increasing the probability that one's beliefs will be unduly influenced by one's desires. Grief may do this.

Do emotions figure more prominently in some cases of self-deception? In the remainder of this essay, I explore two hypotheses about this.<sup>10</sup>

1. The solo emotion hypothesis. In some instances of entering self-deception in acquiring a belief, an emotion, but no desire, makes a biasing contribution to the production of that belief.

2. The direct emotion hypothesis. In some instances of entering self-deception in acquiring a belief, an emotion makes a biasing contribution to the production of that belief that is neither made by a desire nor causally mediated by a desire.<sup>11</sup>

My primary aim is to convey a sense of what may be said for and against these hypotheses and of difficulties involved in investigating them. As I will explain, the second, more modest hypothesis is plausible, and our knowledge about emotion is too thin to warrant a confident rejection or endorsement of the first hypothesis.

Regarding an instance of twisted self-deception, Tim Dalgleish writes: "it is inappropriate to suggest that jealous persons desire or are motivated to find that their partners are unfaithful; rather, their emotional state is priming the relevant processing systems to gather evidence in a biased fashion" (1997, p. 110). Dalgleish's contention is that, in cases of this kind, emotion plays biasing roles of the sort I attributed to desires in straight self-deception. For example, jealousy may prime the confirmation bias by prompting a jealous man to test the hypothesis that his wife is unfaithful, and it may increase the salience of apparent evidence of infidelity. There is evidence that emotions operate in these ways. As Douglas Derryberry reports, there is evidence that "emotional states facilitate the processing of congruent stimuli" and that "attentional processes are involved in [this] effect" (1988, pp. 36, 38), and Gordon

Bower and Joseph Forgas review evidence that emotions make “emotionally congruent interpretations of ambiguous stimuli more available” (2000, p. 106).<sup>12</sup> For example, Jed’s jealousy may make him highly attentive to rare memories of Jane’s seemingly being flirtatious or secretive and help generate jealousy-congruent interpretations of relatively neutral data.

The jealous Jed scenario is unlikely to confirm the solo emotion hypothesis. Sharpsteen and Kirkpatrick, suggest, plausibly, that “the jealousy complex” is “a manifestation of motives reflecting both sexual and attachment concerns” (1997, p. 638). Jealousy is intimately bound up with desires that jealous people have concerning their relationships with the people of whom they are jealous. It is a truism that indifference about one’s relationship with a person precludes being jealous of that person.<sup>13</sup> (Being envious of someone is another matter.) Indeed, it is plausible that if a desire for close romantic attachment is not a constituent of paradigmatic romantic jealousy, it is at least a significant, partial cause of such jealousy. If this plausible proposition is true, then if Jed’s being jealous of Jane affects the hypotheses he frames about her, the vividness of his evidence, and the focus of his attention, it is very likely that an attachment desire plays a biasing role.<sup>14</sup>

Again, “the jealousy complex” can be regarded 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, p. 627). This suggests that the effects of jealousy are partly explained by a desire for the maintenance of a close relationship. That desire may be at work in Jed’s biased cognition. The desire may contribute to Jed’s having a stronger desire to avoid falsely believing that Jane is faithful than to avoid falsely believing that she is unfaithful and, accordingly, contribute to his having a lower acceptance threshold for the hypothesis that she is having an affair than for the contrary hypothesis. The desire, given its psychological context, including, importantly, the jealousy associated with it, may also help enhance the salience of evidence of threats to the maintenance of his relationship with Jane, help prime the confirmation bias in a way favoring the belief that she is having an affair, and so on.

Defending the solo emotion hypothesis is a challenging project. Owing to the tight connection between emotions and associated desires, testing empirically for cases of self-deception in which emotion, and not desire, biases belief is difficult. Constructing compelling conceptual tests also is challenging. For example, if all emotions, or all emotions that might plausibly bias beliefs, are partly constituted by desires, it is difficult to show that there are beliefs that are biased by an emotion, or by some feature of an emotion, but not at all by desires, including desires that are constituents of the biasing emotions. Even if there is a conceptual connection between types of emotions and types of desires as partial causes, rather than between types of emotions and types of desires as constituents, it would have to be shown that emotions sometimes contribute to instances of self-deception to which the desires involved in producing the emotions make no belief-biasing contribution. Furthermore, even if some emotions are neither partially constituted nor partially caused by a relevant desire (typical instances of surprise are like this), the solo emotion hypothesis requires that such an emotion’s biasing contribution to self-deception not be causally mediated by a desire either and, more generally, that the emotion not contribute to self-deception in combination with any biasing desire. I will return to this hypothesis shortly.

The direct emotion hypothesis is more modest.<sup>15</sup> Perhaps in some or many instances of self-deception, biasing roles are played both by (aspects of) emotions and by desires that are intimately related to the biasing emotions – as part to whole, or as a partial cause or effect, or as responses to the emotions (as in the case of a desire to be rid of one’s present anxiety).<sup>16</sup> In some

such cases, some biasing roles played by emotions may be direct, in the relevant sense. Perhaps an emotion can prime the confirmation bias or enhance the salience of emotion-congruent data without doing so simply in virtue of a constituent desire's playing this role and without the effect's being causally mediated by a desire. One who knows only of Jed's evidence for and against the proposition that Jane is having an affair and of his desire for the maintenance of a close relationship with her is hard put to understand why Jed believes that this proposition is true. People with much stronger evidence of infidelity than Jed has often believe that their spouses are innocent of infidelity, even though they, like Jed, strongly desire the maintenance of close relationships with their spouses. Indeed, some common philosophical examples of straight self-deception feature such people. The information that Jed is jealous helps us understand why he believes what he does. His jealousy is an important, instructive part of the psychological context in which he acquires his infidelity belief. Perhaps Jed's jealousy plays a role in the production of his biased belief that is not played by the pertinent desire alone.

Consider another scenario. Ed is angry at Don for a recent offense. His anger may prime the confirmation bias by suggesting an emotion-congruent hypothesis about Don's current behavior – for example, that Don is behaving spitefully again. Ed's anger may also increase the salience of data that seem to support that hypothesis. 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 Ed's anger leads him to view certain aspects of Don's behavior as more goal-directed 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 anger. Possibly, anger can play the biasing roles just mentioned without any constituent desire's playing them and in the absence of causal mediation by a desire.

If an emotion can play a direct biasing role in self-deception, perhaps an emotion may contribute to an instance of self-deception that involves no desires as significant biasing causes. Perhaps the solo emotion hypothesis is true, despite the challenges it faces. It is conceivable, perhaps, that Ed enters self-deception in acquiring the belief that Don is behaving spitefully now, that the process that results in this belief features his anger's playing the biasing roles just described, and that no desires of Ed's have a biasing effect in this case. Now, on the assumption that Ed believes that Don is behaving spitefully despite having stronger evidence for the falsity of that hypothesis than for its truth, an FTL theorist will find it natural to suppose that Ed 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 Ed's acquiring the belief he does. But this supposition is open to debate, and I will not try to settle the issue here.

I mentioned that testing the solo emotion hypothesis empirically would be difficult. This point helps to explain the limited scope that Joseph Forgas claims for his “affect infusion model” of the effects of affective states on social judgments (1995; cf. Bower and Forgas 2000).<sup>17</sup> Sketching some background will enable me to say how. Forgas identifies two “mechanisms of affect-infusion: affect-priming and affect-as-information” (p. 40). The former is a matter of the “selective influence [of affective states] on attention, encoding, retrieval, and associative processes” during substantive information processing (p. 40). In a nice illustration of the latter, “when subjects were asked to make off-the-cuff evaluative judgments about their happiness and life satisfaction through a telephone survey, their responses were significantly different depending on whether they were feeling good (interviewed on a pleasant, sunny day) or feeling bad (interviewed on a rainy, overcast day). Once their attention was called to the source of their

mood (the weather), however, the mood effects were constrained” (p. 53).<sup>18</sup> Commenting on such studies, Norbert Schwarz writes: “rather than computing a judgment on the basis of . . . features of a target, individuals may . . . ask themselves ‘How do I feel about it’ [and] in doing so, they may mistake [certain] feelings . . . as a reaction to the target” (1990, p. 529).

Forgas attempts to demonstrate that “affect infusion is a significant and reliable source of judgmental distortions,” and his model “predicts that affect infusion should not influence judgments based on . . . motivated processing strategies” (p. 51). The “specific goals” he cites as motivators of processing are “mood repair and mood maintenance, self-evaluation maintenance, ego enhancement, achievement motivation, and affiliation” (p. 47; cf. Bower and Forgas 2000, pp. 130-35, 138). Forgas wants to accommodate cases in which, instead of mood-congruent processing, incongruence is found; and his explanation of such cases is partly motivational (cf. Bower and Forgas 2000, pp. 135, 154-55). In mood repair, for example, people selectively attend to memories and thoughts that are incongruent with their unpleasant feelings, motivated by a desire to feel better.

The idea that “goals” such as these are at work in some cases of lay hypothesis testing is easily accommodated by the FTL model. For people experiencing an unpleasant mood or emotion, or a threat to their positive self-image, certain associated errors may be especially costly. For example, in some cases in which people are feeling sad or guilty, the errors of underestimating the quality of their lives or overestimating their responsibility for a harm may be particularly costly. However, this point about the FTL model should not be taken to ground the claim that there is a strict division of labor in lay hypothesis testing between motivation to minimize costly errors and affect infusion. Return to jealous Jed. Owing partly to his jealousy, the most costly error for him may be falsely believing that Jane is faithful, and his processing may be congruent with his jealousy. Similarly, the costliest error for someone who is feeling particularly proud of himself may be falsely believing something that would entail that his pride is unwarranted, and his processing may be congruent with his pride. The question is open whether there is both motivated processing and affect infusion in these scenarios. Forgas apparently commits himself to holding that if motivated processing is at work in them, affect infusion is not. Seemingly, a significant part of what accounts for his taking this view is the difficulty, in such scenarios, of demonstrating empirically that affective states played an infusing role – that, for example, in cases like Jed’s, selective attention to and retrieval of thoughts and images congruent with one’s jealousy is accounted for at least partly by affect infusion rather than solely by other factors, including “motivated processing strategies.”

Again, on Forgas’s model, “judgments based on . . . motivated processing strategies” are not influenced by affect infusion (p. 51). If the FTL theory is correct, all lay hypothesis testing involves motivated processing strategies and Forgas’s claim about his model, literally interpreted, leaves no room for affect infusion in that sphere. Of course, the FTL theory may be overly ambitious, and Forgas may have been overly restrictive in his statement of what his model predicts. However this may be, a method for testing for the joint influence of motivated processing and affect infusion in biased belief would be useful.

A related issue also merits further investigation. Return again to Jed. He wants it to be true that Jane is not having an affair, and he presumably fears at some point that she is. Eventually, he comes to believe that Jane has been unfaithful. Suppose that Jed’s jealousy contributed to that biased belief. Assuming that his jealousy affected his framing of hypotheses, his attention, or the salience of his evidence in a way that contributed to his biased belief that Jane is unfaithful, why didn’t it happen instead that his fear, or a constituent desire, affected



these things in a way that contributed to his acquiring a belief that she is faithful? Alternatively, why didn't his fear, or a constituent desire, block the relevant potential effects of his jealousy, with the result that the balance of his evidence carried the day?<sup>19</sup>

A proponent of the FTL theory might answer these questions in a way that downplays belief-biasing roles for emotions. In a typical case of romantic jealousy where there are some grounds for suspecting infidelity, the belief that one's romantic partner is having an affair would cause psychological discomfort, but it might also promote one's chances of taking successful steps to save one's relationship. It may be suggested (1) that what one believes is determined by a combination of (a) the strength of one's evidence for and against the proposition that one's partner is having an affair and (b) which error one more strongly desires to avoid and (2) that b is determined by the relative strengths of one's desire to avoid the psychological discomfort of believing that one's partner is having an affair and of one's desire to maintain the relationship. However, this view of things may be too simple. Perhaps distinctively emotional features of jealousy can influence what a jealous person believes in a way that does not depend on desire. Furthermore, even if desire and desire-strength are relevant to what a jealous person comes to believe, that is consistent with his jealousy's having a "direct" biasing effect on what he believes.

The questions I raised about Jed are difficult ones.<sup>20</sup> Answers that properly inspire confidence will not, I fear, be produced by philosophical speculation. Nor, as far as I know, are such answers available in the empirical literature on emotion: we need to know more than is currently known about the effects of emotions on cognition. These observations are, of course, consistent both with the truth and with the falsity of the direct and solo emotion hypotheses. I am keeping an open mind and trying to be unbiased.<sup>21</sup>

## REFERENCES

- Barnes, A. (1997). Seeing Through Self-Deception. Cambridge: Cambridge University Press.
- Baron, J. (1988). Thinking and Deciding. Cambridge: Cambridge University Press.
- Bower, G. and Forgas, J. (2000). "Affect, Memory, and Social Cognition." In E. Eich, J. Kihlstrom, G. Bower, J. Forgas, and P. Niedenthal, eds. Cognition and Emotion. Oxford: Oxford University Press.
- Dalgleish, T. (1997). "Once More with Feeling: The Role of Emotion in Self-Deception." Behavioral and Brain Sciences 20, 110-11.
- Davis, W. (1988). "A Causal Theory of Experiential Fear." Canadian Journal of Philosophy 18: 459-83.
- Derryberry, D. (1988). "Emotional Influences on Evaluative Judgments: Roles of Arousal, Attention, and Spreading Activation." Motivation and Emotion 12: 23-55.
- de Sousa, R. (1987). The Rationality of Emotion. Cambridge: MIT Press.
- Forgas, J. (1995). "Mood and Judgment: The Affect Infusion Model." Psychological Bulletin 117: 39-66.
- Friedrich, J. (1993). "Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena." Psychological Review 100: 298-319.
- Gigerenzer, G. and Hug, K. (1992). "Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change." Cognition 43: 127-71.
- Gur, R. and Sackeim, H. (1979). "Self-Deception: A Concept in Search of a Phenomenon." Journal of Personality and Social Psychology 37: 147-69.
- Johnston, M. (1988). "Self-Deception and the Nature of Mind." In B. McLaughlin and A. Rorty, eds. Perspectives on Self-Deception. Berkeley: University of California Press.
- Keltner, D., Ellsworth, P., and Edwards, K. (1993). "Beyond Simple Pessimism: Effects of Sadness and Anger on Social Perception." Journal of Personality and Social Psychology 64: 740-52.
- Klayman, J. and Ha, Y. (1987). "Confirmation, Disconfirmation, and Information in Hypothesis-Testing." Psychological Review 94: 211-28.
- Kunda, Z. (1999). Social Cognition. Cambridge: MIT Press
- \_\_\_\_\_. (1990). "The Case for Motivated Reasoning." Psychological Bulletin 108: 480-98.
- Lazar, A. (1999). "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs 'Under the Influence'." Mind 108: 265-90.
- Mele, A. (2003). Motivation and Agency. New York: Oxford University Press.
- \_\_\_\_\_. (2001). Self-Deception Unmasked. Princeton: Princeton University Press.
- \_\_\_\_\_. (2000). "Self-Deception and Emotion." Consciousness and Emotion 1: 115-37
- \_\_\_\_\_. (1999). "Twisted Self-Deception." Philosophical Psychology 12: 117-37.
- \_\_\_\_\_. (1997a). "Real Self-Deception." Behavioral and Brain Sciences 20: 91-102.
- \_\_\_\_\_. (1997b). "Understanding and Explaining Real Self-Deception." Behavioral and Brain Sciences 20: 127\_34.
- \_\_\_\_\_. (1987). Irrationality. New York: Oxford University Press.
- Nisbett, R. and Ross, L. (1980). Human Inference: Strategies and Shortcomings of Social Judgment. Englewood Cliffs: Prentice-Hall.
- Pears, D. (1984). Motivated Irrationality. Oxford: Oxford University Press.
- Quattrone, G. and Tversky, A. (1984). "Causal Versus Diagnostic Contingencies: On Self-Deception and on the Voter's Illusion." Journal of Personality and Social Psychology 46: 237-

48.

Schwarz, N. (1990). "Feelings as Information: Informational and Motivational Functions of Affective States." In E. Higgins and R. Sorrentino, eds. Handbook of Motivation and Cognition, vol. 2. New York: Guilford Press.

Schwarz, N. and Clore, G. (1983). "Mood, Misattribution and Judgments of Well-Being." Journal of Personality and Social Psychology 45: 513-23.

Sharpsteen, D. and Kirkpatrick, L. (1997). "Romantic Jealousy and Adult Romantic Attachment." Journal of Personality and Social Psychology 72: 627-40.

Tesser, A., Pilkington, C., and McIntosh, W. (1989). "Self-Evaluation Maintenance and the Medial Role of Emotion: The Perception of Friends and Strangers." Journal of Personality and Social Psychology 57: 442-56.

Trivers, R. (1985). Social Evolution. Menlo Park, CA: Benjamin/Cummings.

Trope, Y., Gervy, B., and Liberman, N. (1997). "Wishful Thinking from a Pragmatic Hypothesis-Testing Perspective." In M. Myslobodsky, ed. The Mythomanias: The Nature of Deception and Self-Deception. Mahwah, NJ: Lawrence Erlbaum.

Trope, Y. and Liberman, A. (1996). "Social Hypothesis Testing: Cognitive and Motivational Mechanisms." In E. Higgins and A. Kruglanski, eds. Social Psychology: Handbook of Basic Principles. New York: Guilford Press.

## NOTES

<sup>1</sup> This tradition is embraced in influential work on self-deception in philosophy, psychology, psychiatry, and biology. See, e.g., Pears 1984, Quattrone and Tversky 1984, Gur and Sackeim 1979, and Trivers 1985. Stereotypical interpersonal deception does not exhaust interpersonal deception.

<sup>2</sup> Two points should be made. First, I have never defended a statement of necessary and sufficient conditions of entering self-deception in acquiring a belief that p, but only statements of characteristic and jointly sufficient conditions. (For a recent statement, see Mele 2001, pp. 50-51.) Second, 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>3</sup> On deviant and nondeviant causation in this connection, see Mele 2001, pp. 121-23.

<sup>4</sup> Cf. Mele 2001, p. 106. The improvement is the "reflection" clause. An issue may be so boring to one's impartial cognitive peers that they do not reflect on it and reach no conclusion about it.

<sup>5</sup> I develop this idea in Mele 1987, ch. 10 and Mele 2001. Kunda 1990 develops the same theme, concentrating on evidence that motivation sometimes primes the confirmation bias. Also see Kunda 1999, ch. 6.

<sup>6</sup> For motivational interpretations of the confirmation bias, see Friedrich 1993 and Trope and Liberman 1996, pp. 252-65.

<sup>7</sup> See Mele 2001, pp. 31-49, 63-70, 90-91, 96-98, 112-18.

<sup>8</sup> On this case, see Barnes 1997, ch. 3; Dalgleish 1997, p. 110; Lazar 1999, pp. 274-77; and Pears 1984, pp. 42-44.

Also see Mele 1987, pp. 114-18.

<sup>9</sup> Fear that  $\sim p$  is plausibly understood as being partly constituted by desire that p. See, e.g., Davis 1988.

<sup>10</sup> Elsewhere (Mele 2000, pp. 125-29), I criticized a third hypothesis, what I called "the anxiety reduction hypothesis" – the thesis that the function of self-deception is to reduce present anxiety (Barnes 1997, Johnston 1988; cf. Tesser *et al.* 1989).

<sup>11</sup> To simplify discussion, I formulated both hypotheses in terms of entering self-deception in acquiring a belief.

Entering self-deception in retaining a belief and remaining in self-deception in continuing to believe something also require attention.

<sup>12</sup> Reviews of the "mood congruence effect" include Bower and Forgas 2000 and Forgas 1995.

<sup>13</sup> If a woman is jealous because her date is flirting with another woman, is she jealous of her date or the other woman? Ronald de Sousa expresses the proper usage succinctly: "the person one is jealous of plays an entirely different part in one's jealousy from that of the rival because of whom one is jealous" (1987, p. 75).

<sup>14</sup> The conjunction of “x affects y” and “z is a constituent or a cause of x” does not entail “z affects y.” The brake pedal on Smith’s car is a constituent of his car and his car affected Jones. But the brake pedal did not. Smith’s car fell on Jones as he was repairing a flat tire. That explains the qualification “very likely” in the text.

<sup>15</sup> More specifically, although, necessarily, any emotion that makes a solo biasing contribution to self-deception makes a direct biasing contribution (in the pertinent sense of ‘direct’), an emotion that makes a direct biasing contribution might not make a solo one.

<sup>16</sup> The categories of effect and response are not mutually exclusive.

<sup>17</sup> Forgas uses ‘affect’ as “a generic label to refer to both moods and emotions” (p. 41).

<sup>18</sup> This experiment is reported in Schwarz and Clore 1983.

<sup>19</sup> It may be suggested that Jed’s fear issued in fear-congruent processing that meshed with his jealousy-congruent processing. Even if that is so, one wants to understand why the desire-component of his fear – his desire that Jane is not having an affair – did not contribute to motivated processing resulting in a belief that she is innocent of infidelity, or block effective jealousy-congruent processing.

<sup>20</sup> A proponent of the direct emotion hypothesis may urge that occurrent aversions to specific costly mistakes do the work attributed to desires to avoid these mistakes in my sketch of the FTL theory, that these aversions play the role attributed to the avoidance desires in determining confidence thresholds. Such a theorist may also contend that occurrent aversions are emotions and argue that even though an aversion to falsely believing that p has a desire to avoid falsely believing that p as a constituent, the work of the aversion in biasing belief is not exhausted by the biasing work of the desire. The claim may be that some distinct, affective feature of the aversion makes a biasing contribution of its own to confidence thresholds. It may also be claimed that the aversion plays additional direct roles – for example, enhancing the salience of evidence for  $\sim p$ . This is another issue that requires further investigation, including conceptual spadework. Once it is suggested that occurrent aversions are emotions, the suggestion that all desires – or all desires with some felt intensity – are emotions may not be far behind.

<sup>21</sup> Parts of this article derive from Mele 1987, 1997a, 2000, 2001, and 2003. I am grateful to audiences at the University of Zurich Ethics Center and the University of Manchester for fruitful discussion.