

When Are We Self-Deceived?*

Alfred R. Mele[†]
almele@fsu.edu

ABSTRACT

This article's point of departure is a proto-analysis that I have suggested of entering self-deception in acquiring a belief and an associated set of jointly sufficient conditions for self-deception that I have proposed. Partly with the aim of fleshing out an important member of the proposed set of conditions, I provide a sketch of my view about how self-deception happens. I then return to the proposed set of jointly sufficient conditions and offer a pair of amendments.

Introduction

In *Self-Deception Unmasked* (Mele 2001) and in earlier work, I tried to show that self-deception is masked by traditional models of the phenomenon that treat it as an intrapersonal analogue of stereotypical interpersonal deception.¹ According to these models, self-deceivers intentionally deceive themselves into believing that p , and there is a time at which they believe that p is false while also believing that p is true. In Mele 2001, I offered an alternative model of self-deception and, drawing heavily on empirical work, I developed a detailed explanation of how garden-variety self-deception happens.

The contributors to this issue have been asked to focus on philosophical aspects of self-deception. I focus here on a question about conceptually sufficient conditions for self-deception. In section 1, I review a proto-analysis that I have suggested of entering self-deception in acquiring a belief and an

* In parts of this article, I draw on Mele 2001 and 2009. This article was made possible through the support of a grant from the John Templeton Foundation. The opinions expressed in this article are my own and do not necessarily reflect the views of the John Templeton Foundation. I am grateful to an anonymous referee for comments on a draft of this article.

[†] Florida State University, USA.

¹ 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.

associated set of jointly sufficient conditions for self-deception that I have proposed. In section 2, partly with the aim of fleshing out an important member of the proposed set of conditions, I provide a sketch of my view about how self-deception happens. In section 3, I return to the proposed set of jointly sufficient conditions and offer two amendments.

1. A Proto-Analysis and Proposed Sufficient Conditions

Although I have never offered a conceptual analysis of self-deception, I have suggested the following proto-analysis of entering self-deception in acquiring a belief: 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). 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 . My suggestion is that someone interested in constructing a conceptual analysis of entering self-deception in acquiring a belief can start here and try to work out an account of suitable bias. Of course, an analysis of entering self-deception in acquiring a belief will not be a complete analysis of self-deception if there are other ways of entering self-deception; and, as I have explained elsewhere, people sometimes enter self-deception in *retaining* a belief (Mele, 2001, pp. 56-59). Someone who faultlessly acquires the belief that p may later enter self-deception in persisting in believing that p . It may be suggested that if a complete analysis of self-deception is constructable, it is constructable out of analyses of these two ways of entering self-deception.²

I have also proposed a set of conceptually sufficient conditions for self-deception, as follows:

Senters self-deception in acquiring a belief that p if:

1. The belief that p which S acquires is false,
2. S treats data relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way,
3. This biased treatment is a nondeviant cause of S 's acquiring the belief

² Some theorists would definitely reject this suggestion. Robert Audi, for example, contends that no one who is self-deceived about p has a false belief that p ; rather, self-deceived people have an unconscious true belief that $\sim p$ and – in the absence of a belief that p – sincerely avow that p (1982, 1985, 1997). I criticize Audi's attempted analysis of self-deception in Mele 1982 and 2010; I will not do so again here.

- that p , and
4. The body of data possessed by S at the time provides greater warrant for $\sim p$ than for p . (Mele 2001, pp. 50-51; see Mele 1997, p. 95)

I comment briefly on each condition and then forge ahead.

Condition 1 captures a purely lexical point. A person is, by definition, *deceived in* believing that p only if p is *false*; the same is true of being *self-deceived in* believing that p . The condition does not imply that the falsity of p has special importance for the *dynamics* of self-deception. Motivationally 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 .

People may be deceived *into* believing something that they are not deceived *in* believing (see Mele 1987, pp. 127-28). Ann might execute a complicated strategy for deceiving Alan into believing something that, unbeknownst to her, is true. And she might thereby cause him to believe this proposition, p . Since p is true, Alan is not deceived *in* believing it. Even so, it is plausible that Ann deceived him *into* believing it, if she caused him to believe that p partly by deceiving him into believing some false propositions suggestive of p .

My discussion of motivated bias and various ways of entering self-deception in the following section puts some flesh on the bones of condition 2. An interpretation of condition 2 will emerge from that section.

My inclusion of the term "nondeviant" in condition 3 is motivated by a familiar problem for causal characterizations of phenomena in any sphere. Specifying the precise nature of nondeviant causation of a belief by motivationally biased treatment of data is a difficult technical task. Mele 2001 provides guidance on the issue.

The thrust of condition 4 is that self-deceivers believe against the weight of the evidence they possess. I do not view 4 as a *necessary* condition of self-deception. In some instances of motivationally biased evidence-gathering, for example, people may bring it about that they believe a falsehood, p , when $\sim p$ is much better supported by evidence readily available to them, even though, owing to the selectivity of the evidence-gathering process, the evidence that they themselves actually *possess* at the time favors p over $\sim p$. In my view, such people are naturally deemed self-deceived, other things being equal. However, some philosophers require that a condition like 4 be satisfied (Davidson 1985, McLaughlin 1988, Szabados 1985), and I do not object to including 4 in a list

of jointly *sufficient* conditions. Of course, in some cases, whether the weight of a person's evidence lies on the side of p or of $\sim p$ (or equally supports each) is subject to legitimate disagreement.

2. Explaining 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 spouse is not having an affair. 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 wanting it to be false that she is so engaged (and not also wanting it to be true that she is). In cases of both kinds, as I have explained in Mele 2001 and briefly explain below, self-deceivers have motivationally biased beliefs.

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 hypersensitivity 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, if we assume that these people acquire relevant false, unwarranted beliefs in the

ways described, these are garden-variety instances of self-deception; or so I have argued elsewhere.³ 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 (Mele 2001, pp. 28–31), including the following three:

(a) *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 (Nisbett and Ross 1980).

(b) *The availability heuristic*. When we form beliefs about the frequency, likelihood, or causes of an event, we «often may be influenced by the relative availability of the objects or events, that is, their accessibility in the processes of perception, memory, or construction from imagination» (Nisbett and Ross, 1980, p. 18). For example, we may mistakenly believe that the number of English words beginning with 'r' greatly outstrips the number having 'r' in the third position, because we find it much easier to produce words on the basis of a search for their first letter (Tversky & Kahnemann, 1973). Similarly, attempts to locate the cause(s) of an event are significantly influenced by manipulations that focus one's attention on a specific potential cause (Nisbett and Ross, 1980, p. 22; Taylor & Fiske, 1975, 1978).

³ 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.

(c) *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. 259–265). 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, Gervy, & Liberman, 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.⁴ 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; and vivid or salient data, given that they are more likely to be recognized and recalled, tend to be more "available" than pallid counterparts. Similarly, desires can influence which hypotheses occur to one and affect the salience of available hypotheses, thereby setting the stage for the confirmation bias.⁵ 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 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.⁶ I call it the *FTL theory*, after the authors of the two articles on which I primarily drew, Friedrich 1993 and Trope & Liberman 1996. Here, I offer a sketch of the theory.

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" –

⁴ 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.

⁵ For motivational interpretations of the confirmation bias, see Friedrich 1993 and Trope and Liberman 1996, pp. 252–265.

⁶ See Mele 2001, pp. 31–49, 63–70, 90–91, 96–98, 112–18.

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 the test» (Trope & 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» (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).⁷

Friedrich claims that desires to avoid specific errors can trigger and sustain «automatic test strategies» (1993, p. 313), which 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

⁷ 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 & Hug 1992.

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 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. (1993, p. 314)

Here, he 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» (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).

3. Proposed Sufficient Conditions Revisited

I return to my proposed set of jointly sufficient conditions for entering self-deception in acquiring a belief. Some philosophers have argued that my four conditions fall short of collective sufficiency because they do not capture a kind of *tension* that is necessary for self-deception. According to Robert Audi, this tension «is ordinarily represented [...] by an avowal of p [...] *coexisting* with knowledge or at least true belief that $\sim p$ » (1997, p. 104). Eric Funkhouser claims that self-deception requires tension between some of the agent's behavior and certain of her sincere avowals (2005, p. 304). Michael Losonsky contends that self-deceivers have the unwarranted, false belief that p , lack the true belief that $\sim p$, and have evidence for $\sim p$ that is «active» in their «cognitive architecture» (1997, p. 122). This activity, he claims, is manifested in such indications of tension as recurrent or nagging doubts, and he uses the contention that self-deception conceptually requires such conflict to support a distinction between self-deception and instances of “prejudice” or “bias” that satisfy the quartet of conditions I offered as conceptually sufficient for entering self-deception. Mike W. Martin mentions a similar tension, «a cognitive conflict» such as «suspecting p and believing $\sim p$ » (1997, p. 123). And Kent Bach maintains that self-deception requires actively avoiding or suppressing certain thoughts, or ridding oneself of these thoughts when they occur (1997; also see Bach 1998, pp. 167–168).

The quartet of conditions I offered certainly does not entail that there is no tension in self-deception. Nor do I claim that self-deception normally is tension-free. Significant tension may be present in most people who satisfy my four conditions. But the issue raised by the authors mentioned in the preceding paragraph is whether the alleged kinds of tension are conceptually *necessary* for entering self-deception. And my answer has been *no*. As I see it, given the details of Rex's story, even if he is tension-free during the process of acquiring the belief that his article was wrongly rejected and while that belief is in place, he is self-deceived and he enters self-deception in acquiring that

belief. In my view, the same is true of bigots who, without psychic conflict, satisfy my four conditions in acquiring a bigoted belief that p .

The primary topic of the present section is conceptually sufficient conditions for entering self-deception in acquiring a belief – not individually necessary and jointly sufficient conditions for this. Different philosophers require different kinds of tension for self-deception, as the first paragraph of this section suggests; and I have argued that tension of the various kinds at issue is not required for self-deception (Mele 2001). But even if I am right in keeping tension off a list of necessary conditions of self-deception, it may appear on a useful list of jointly sufficient conditions. The following addition to my proposed quartet of jointly sufficient conditions (which resembles Martin’s condition of suspecting that the pertinent proposition one believes is false [1997, p. 123]) would result in a less latitudinarian proposal about sufficient conditions for entering self-deception: (5) S consciously believes at the time that there is a significant chance that $\sim p$ (see Mele 2001, pp. 71–73 and 2010, p. 749). For example, the resulting proposal would not entail that tension-free Rex enters self-deception in acquiring the belief that his submission was wrongly rejected.

The second and third conditions in my proposed set of sufficient conditions include the expressions “ S treats data” and “This biased treatment.” I intended my discussion (in Mele 1997 and 2001) of various ways of entering self-deception in acquiring a belief that p to provide guidance on the interpretation of “treats” and “treatment” in these conditions. But if, strictly speaking, relatively simple motivationally biased misperception counts as motivationally biased *treatment* of data (given the standard meaning of “treats data”), trouble is brewing. Imagine that a hungry cat misperceives a noise as the sound of her food being shaken into a bowl and runs into the room from which the noise is emanating (Scott-Kakures 2002, pp. 578–580). Those who are happy to attribute beliefs to cats may be happy to say that the cat has a belief to the effect that food is available, and that belief may be a relatively direct product or a constituent of her motivationally biased misperception of the noise. If feline self-deception is out of the question and if “treats data” has a broader sense than I intended, then something should be done about “treats” in condition 2 or a useful condition should be added. How should this be handled?

Dion Scott-Kakures argues that «reflective, critical reasoning is essential to the process of self-deception» (2002, p. 577) and that «the error of self-

knowledge that makes [...] self-deception possible is a misconception about what animates [the believer's] doxastic or cognitive activities. Like any reflective reasoner, she will regard her investigations as directed by [...] her grasp upon what reason recommends,» but she is wrong about this (p. 599). «Her investigations are directionally driven by desire or interest» (p. 599), in ways featured in my account of how self-deception happens. If Scott-Kakures is right in requiring these things for self-deception, something like the following condition should be added to my proposed sufficient conditions for S 's entering self-deception in acquiring a belief that p : (6) S 's acquiring the belief that p is a product of “reflective, critical reasoning,” and S is wrong in regarding that reasoning as properly directed.⁸ I have no objection to including condition 6 in a list of jointly sufficient conditions for entering self-deception in acquiring a belief that p .

Putting things together, I arrive at the following statement of proposed jointly sufficient conditions for entering self-deception in acquiring a belief:

S enters self-deception in acquiring a belief that p if:

1. The belief that p which S acquires is false
2. S treats data relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way
3. This biased treatment is a nondeviant cause of S 's acquiring the belief that p
4. The body of data possessed by S at the time provides greater warrant for $\sim p$ than for p
5. S consciously believes at the time that there is a significant chance that $\sim p$
6. S 's acquiring the belief that p is a product of “reflective, critical reasoning,” and S is wrong in regarding that reasoning as properly directed.

My primary aim in previous work on self-deception has been to explain how it happens. The explanation I developed elsewhere and sketched in section 2 applies straightforwardly to cases in which these six conditions are satisfied.

⁸ Scott-Kakures motivates a condition of this kind not only by means of reflection on the case of the hungry cat, but also by means of reflection on «“precipitate cases” of motivated believing» in human beings (2002, p. 587), cases in which a person leaps to a motivationally biased conclusion in the absence of reflective reasoning.

REFERENCES

- Audi, R. (1982). Believing and Affirming. *Mind*, *91*(361), 115–120.
- Audi, R. (1985). Self Deception and Rationality. In M. Martin (Ed.), *Self Deception and Self Understanding*. Lawrence: University of Kansas Press, 169–194.
- Audi, R. (1997). Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele. *Behavioral and Brain Sciences*, *20*(1), 104.
- Bach, K. (1997). Thinking and Believing in Self-Deception. *Behavioral and Brain Sciences*, *20*(1), 105.
- Bach, K. (1998). (Apparent) Paradoxes of Self-Deception and Decision. In J. Dupuy (Ed.), *Self-Deception and Paradoxes of Rationality*. Cambridge: Cambridge University Press, 163–189.
- Baron, J. (1988). *Thinking and Deciding*. Cambridge: Cambridge University Press.
- Davidson, D. (1985). Deception and Division. In E. LePore & B. McLaughlin (Eds.), *Actions and Events*. Oxford: Basil Blackwell, 138–148.
- Friedrich, J. (1993). Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena. *Psychological Review*, *100*(2), 298–319.
- Funkhouser, E. (2005). Do the Self-Deceived Get what They Want. *Pacific Philosophical Quarterly*, *86*(3), 295–312.
- Gigerenzer, G., & Hug, K. (1992). Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change. *Cognition*, *43*, 127–171.
- Kunda, Z. (1990). The Case for Motivated Reasoning. *Psychological Bulletin*, *108*(3), 480–498.
- Kunda, Z. (1999). *Social Cognition*. Cambridge, Mass.: MIT Press.
- Losonsky, M. (1997). Self-Deceivers' Intentions and Possessions. *Behavioral and Brain Sciences*, *20*(1), 121–122.

- Martin, M. (1997). Self-Deceiving Intentions. *Behavioral and Brain Sciences*, 20(1), 122–123.
- McLaughlin, B. (1988b). Exploring the Possibility of Self-Deception. In B.P. McLaughlin & A.O. Rorty (Eds.), *Perspectives on Self-Deception*. Berkeley: University of California Press, 1988, 29–62.
- Mele, A.R. (1982). Self-Deception, Action, and Will: Comments. *Erkenntnis*, 18(2), 159–164.
- Mele, A.R. (1987). *Irrationality*. New York: Oxford University Press.
- Mele, A.R. (1997). Real Self-Deception. *Behavioral and Brain Sciences* 20(1), 91–102.
- Mele, A.R. (1999). Twisted Self-Deception. *Philosophical Psychology*, 12(2), 117–137.
- Mele, A.R. (2001). *Self-Deception Unmasked*. Princeton: Princeton University Press.
- Mele, A.R. (2009). Have I Unmasked Self-Deception or Am I Self-Deceived? In C. Martin (Ed.), *The Philosophy of Deception*. New York: Oxford University Press, 260–276.
- Mele, A.R. (2010). Approaching Self-Deception: How Robert Audi and I Part Company. *Consciousness and Cognition*, 19(3), 745–750.
- Nisbett, R., & Ross, L. (1980). *Human Inference: Strategies and Shortcomings of Social Judgment*. Englewood Cliffs: Prentice-Hall.
- Scott-Kakures, D. (2002). At ‘Permanent Risk’: Reasoning and Self-Knowledge in Self-Deception. *Philosophy and Phenomenological Research*, 65(3), 576–603.
- Sharpsteen, D, & Kirkpatrick, L. (1997). Romantic Jealousy and Adult Romantic Attachment. *Journal of Personality and Social Psychology*, 72(3), 627–640.
- Szabados, B. (1985). The Self, its Passions, and Self-Deception. In M. Martin (Ed.), *Self-Deception and Self-Understanding*. Lawrence: University of Kansas Press, 143–168.

- Taylor, S., & Fiske, S. (1975). Point of View and Perceptions of Causality. *Journal of Personality and Social Psychology*, 32(3), 439–445.
- Taylor, S., & Fiske, S. (1978). Saliency, Attention and Attribution: Top of the Head Phenomena. In Leonard Berkowitz (Ed.), *Advances in Experimental Social Psychology*, vol. 11. New York: Academic Press, 249-288.
- Trope, Y., & Liberman, A. (1996). Social Hypothesis Testing: Cognitive and Motivational Mechanisms. In E. Higgins & A.W. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles*. New York: Guilford Press, 239–270.
- Trope, Y., Gervy, B., & 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, 105–131.
- Tversky, A., & Kahnemann, D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5(2), 207–232.

