

The Motivating Influence of Emotion on Twisted Self-Deception

Mario R. Echano

Abstract: The question on whether self-deception is intentional or not has divided philosophers into two conflicting sides. Despite the disagreement, partisans of either side tend to converge on characterizing self-deception as a kind of motivated believing. They generally agree that self-deception is motivated by desire. In fact, the basis by which they classify cases of self-deception as straight or twisted is on how desire influences the acquisition of self-deceptive belief. In the former, the desire that p (or the desire to believe that p) influences the subject's acquisition of a belief that p . In the latter, despite not desiring that p to be the case, S still acquires the belief that p . Twisted cases of self-deception, however, pose themselves as challenge to the claim that self-deception is motivated by desire. They are problematic because desiring something undesirable is a contradiction. Taking the nonintentional side of the debate, I aim to explore the most viable explanation on how motivation works on self-deception. I argue that emotions are as responsible as desire in self-deceptive belief acquisition. Following the model of lay-hypothesis testing originally laid out by social psychologists, the self-deceiver is considered as someone testing her hypothesis for its confirmation rather than for its negation. On this model, the role of desire and emotions in self-deception can be seen in the generation of the hypothesis and its actual testing. The motivating influence of emotions in biased belief acquisitions is more obvious in twisted cases especially in the triggering of the hypothesis, whereas desire's influence dominates the triggering of a hypothesis in the straight ones.

Keywords: motivation, cognitive biases, motivated believing, lay-hypothesis testing theory

1. Introduction

The debate on whether self-deception (“SD”, from hereon) is intentional or not makes it difficult for theorists to agree on a definition. Most of those who favor that SD is intentional equate it with lying to oneself, while those who claim otherwise prefer to consider it as either nonintentional misleading or as possession of motivationally biased belief. Despite the above disagreement, both intentionalists and nonintentionalists coincide on the claim that SD is motivated, according to which desire or emotions have a lot to do with its acquisition.

The motivating influence of desire in SD is the basis of one of the ways by which philosophers classify its vast and various cases. They divide SD cases based on how desire exerts an influence in the acquisition of SD belief, namely, as straight and twisted. In the former, the desire that p (or the desire to believe that p) influences the subject’s acquisition of a belief that p (“ S ” refers to the subject, while “ p ” or “ q ” to the proposition that is believed or desired). In the latter, despite not desiring p to be the case, S still acquires the belief that p .

Examples of the straight cases are numerous: in spite of overwhelming evidence of her husband’s infidelity, Laura still believes that he is faithful; Sid has been pursuing Mary for years, but despite being rejected several times he still believes that his love is reciprocated; the emperor in Andersen’s tale (“The Emperor’s New Suit”) believes that he is wearing a unique dress even though it is clear to him that he is naked.¹ The main idea about desires and their role in SD is the following: because of the desire for those beliefs to be true, self-deceivers fail to recognize the available evidence contrary to their beliefs.

In the twisted cases, the object of SD is an undesirable belief: a jealous husband acquires a false belief that his wife is unfaithful despite not wanting her to be so; anorexic Trisha falsely believes that she has a plump body even though she is thin; or Sylvia, who without wanting that she left the gas stove on, ends up believing that she left it on when in reality she did not. These cases are problematic since they pose a challenge as to how they can fit within the desire-based accounts of both intentionalists and the nonintentionalists. The question arises as to how one can desire the undesirable belief.

In the light of the twisted cases, the role of desire in the process of acquisition of SD beliefs become kind of mysterious. The self-deceivers in the above cases seem to acquire a highly undesirable belief. It is apparent that in them, desire is not fit to give a satisfactory explanation and so emotions are

¹ Hans Christian Andersen, “The Emperor’s New Suit,” in *Fairy tales of Hans Christian Andersen* (Auckland: Floating Press, 2014), 234–240.

to be called upon. But if we posit an explanation apart from desire's influence, we might also be forced to admit, like some theorists (e.g. Gardiner, Lazar, etc.), that indeed there is no homogenous explanation for SD and that each case must be treated as unique.² If it were so, it would be harder to identify what really counts as SD.

Other theorists (e.g. Pears, Dalglish, Mele) invoke the role of emotions for twisted SD.³ Indeed, Pears regards such instances of twisted SD as emotional cases.⁴ However, how exactly emotions work along the process of SD belief acquisition is still a matter of controversy. It becomes more problematic because some theorists equate motivation to desire and thus a line is drawn between the concepts of emotions and motivation. On the one hand, Dalglish, Lazar, and Mele speak of an emotional biasing influence different from motivational ones to refer to the biasing influence caused by desire. On the other hand, for most theorists (e.g. Scott-Kakures, Barnes, etc.), there is no such distinction.⁵ They are simply motivational states responsible for the SD belief acquisition.

In any case, the above distinction and the seeming impossibility to provide a unified approach to SD complicate the assessment of the role of emotions in SD. And so, even if desires and emotions can be both understood as motivations, there is a need to clarify their respective roles in the process of motivationally biased belief acquisition.

In this work, I aim to explore the role of emotions in SD, thereby exposing that both cases of SD undergo the same processes of biased belief acquisition. While this role is not apparent in those desirable cases, this role is more tangible in those undesirable ones. By adopting the model of lay-hypothesis testing originally introduced in social psychology in explaining the process of motivated belief acquisition, a homogenous explanation can be provided. In section 2, I will situate the problem within the debate. I will side with the nonintentionalists in their claim that SD is not necessarily intentional. In the third and fourth sections, I will consider how some intentional and nonintentional accounts have dealt with the problem of the

² See Patrick Gardiner, "Error, Faith, and Self-Deception," in *Proceedings of the Aristotelian Society* 70 (1970), 221–240; and Ariela Lazar, "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs Under the Influence," in *Mind* 108:430 (1999), 265–290.

³ See David F. Pears, *Motivated Irrationality* (New York: Oxford University Press, 1984), 41–50; Tim Dalglish, "Once More with Feeling: The Role of Emotion in Self-Deception," in *Behavioral and Brain Sciences* 20 (1997), 110–111; and Alfred Mele, "Emotion and Desire in Self-Deception," in *Royal Institute of Philosophy Supplement*, ed., by Anthony Hatzimoysis (Cambridge: Cambridge University Press, 2003), 163–179.

⁴ Pears, *Motivated Irrationality* 43–44.

⁵ See Dion Scott-Kakures, "Motivated Believing: Wishful and Unwelcome," in *Nous*, 34:3 (2000), 348–375; and Annette Barnes, *Seeing Through Self-deception* (United Kingdom: Cambridge University Press, 1997).

twisted cases of SD. In the last section, I will attempt to sketch how emotions can trigger the processes for the cognitive endorsement of an undesirable belief which can thereby activate the acquisition of that motivationally biased belief.

2. Intentional and NonIntentional Self-Deception Debate

The intentionalists and the nonintentionalists have different ways of assessing cases of SD such as those examples above. Intentionalists often interpret them as modeled after deception of others. More colloquially, they speak of the self-deceivers as lying to themselves. For them, the self-deceivers, while believing that p , intend to make themselves believe that not- p . The emperor, for example, believes that he is naked while he tries to make himself believe that he has a wonderful suit on; or the jealous husband believing that his wife is faithful lies to himself when he believes that she is unfaithful.

But treating those cases of SD as a kind of lying to oneself is problematic. The problem will be clearer if we begin by taking a usual case of lying to others as an example: the case of my lying to Antonio for instance. When I lie to him that tomorrow is my birthday, my deceptive intention can be fulfilled only if he is not aware of my intention. Otherwise, I will not be able to deceive him. Moreover, if my lie succeeds, I believe that not- p while he believes that p (where p is “tomorrow is my birthday”). Once this scenario is applied to “lying to myself,” the difficulty becomes obvious. If I am going to lie to myself that p , I must not let myself know that I intend (or plan or try) to deceive myself, or else I won’t succeed. Also, if we accept the analogy between other-deception and SD, the self-deceiver will hold two contradictory beliefs, i.e., that p and that not- p . In other words, I would believe that it is my birthday tomorrow and that it is not my birthday tomorrow. Baghramian and Nicholson characterize the two conditions for SD modeled after lying to oneself as:

- A) Dual-belief condition: the self-deceived subject simultaneously holds (at least at one time point) two contradictory beliefs: p and not- p .
- B) Deceptive intention condition: the subject intends or tries to deceive herself.⁶

⁶ See Maria Baghramian and Anna Nicholson, “The Puzzle of Self-Deception,” in *Philosophy Compass* 8:11 (2013), 1018.

Mele effectively speaks of two fatal paradoxes which are the results of these conditions as: the *dynamic* or intention paradox, consisting in the difficulty to imagine how the self-deceiver can succeed in deceiving herself when she already knows what she is up to; and the *static* paradox, which is about the psychologically questionable state of the subject's holding of two contradictory beliefs.⁷

Intentionalists have offered several solutions to overcome the puzzles. The most common strategy is by introducing a certain partition within the self which could be of three types. The most moderate will be those of Demos, McLaughlin, and Bermudez, who in one way or another suggest that *S* believes that *p* and not-*p* at the same time while not being aware that *S* believes so.⁸ The most extreme partitioning strategy will be that of King-Farlow and Rorty who introduced several selves within the *S*, allowing each of them to be deceiver and deceived at different turns.⁹ At the middle will be those of Davidson whose mental partitioning allows *S* to believe that *p* because of her belief that not-*p*;¹⁰ and Pears whose division between a main system and a subsystem within *S* allows for the possibility of the subsystem to intentionally deceive the main system, and thus, *S* believes that *p* and that not-*p* at the same time.¹¹ Partitioning strategies apparently solve both the *dynamic* and the *static* paradoxes. It solves the former because it allows different centers of agencies within the self who are capable of deceptive intention. It also solves the latter because these different centers of agency within the self are also capable of holding beliefs that are contradictory.

However, these solutions have met a lot of criticisms because they generate a set of puzzles more problematic than the initial paradoxes. For example, the degrees of autonomy and intentionality attributed to the subsystems for them to be able to deceive each other have led to the problem of infinite regress. Sissela Bok says that if we postulate that the selves are themselves split into selves capable of deceiving one another, we may end up with a myriad of self-propagating little self.¹² Another objection is that this

⁷ See Alfred Mele, *Self-deception Unmasked* (United States: Princeton University Press, 2001), 59–67.

⁸ See Raphael Demos, "Lying to Oneself," in *The Journal of Philosophy* 57:18 (1960); Brian P. McLaughlin, "Exploring the Possibility of Self-Deception in Belief," in *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988), 29–62; and José Luis Bermúdez, "Self-Deception, Intentions, and Contradictory Beliefs," in *Analysis* 60:4 (2000), 309–319.

⁹ See John King-Farlow, "Self-deceivers and Sartrean Seducers," in *Analysis* 23 (1963), 131–136; and Amelie Rorty, "The Deceptive Self: Liars, Layers, and Lairs," in *Perspectives on Self-deception*, 11–28.

¹⁰ See Donald Davidson, "Deception and Division," in *The Multiple Self*, ed. by Jon Elster (Cambridge: Cambridge University Press, 1985).

¹¹ See David F. Pears, "The Goals and Strategies of Self-Deception," in *The Multiple Self*.

¹² See Sissela Bok, "The self deceived," *Social Science Information* 19:6 (1980), 931.

solution to divide the self into subsystems is just another type of interpersonal deception, only that the partitioner has simply substituted interhumoncular deception for SD.¹³

Other intentionalists suggested a different strategy to solve the paradox. They introduce the temporal partitioning (time slicing strategy) that leads to self-induced deception. An example of this case would be Marta who wants to forget about a meeting fixed in a month. So that she may miss it, she writes a wrong date on her diary. Given her poor memory, she trusts that in a month she will believe her own writing and forget about the original date. She then believes the false date and disbelieves the factual date. This strategy, however, has also been criticized for not being a case of SD. What Marta did was to put herself in the condition of believing *p*. Besides, no dual believing really happened. She does not believe that there was a meeting until she learned so later. And by time she realized she did miss the meeting, she now believes only that there was a meeting. This case may be intentional but there is no possession of contradictory beliefs.

On these grounds, nonintentionalists found the intentionalists' accounts of SD unsatisfactory. Since deceptive intention is what makes SD puzzling, nonintentionalists denied that it is necessary for SD. Common among the nonintentionalists' strategy is to deflate the dual belief requirement to possession of only a false belief and the intentional requirement to motivational influences (e.g. desire and emotions) in acquiring a biased belief. For them, it is not necessary for the self-deceiver to intend to deceive themselves. In the case of the emperor for example, without such deceptive intention, he fell into believing that he is wearing a new suit. And although he may be aware that he is naked, he just believes that he is fully clothed. The same interpretation could be given to other cases. The jealous husband does not really have any intention to deceive himself about his wife's infidelity; he just found himself so deceived.

But there are also sound objections against the nonintentional accounts. I will point out three of the most basic. First, because of their deflationary approach, they seem to be talking about a phenomenon other than SD. There is a tendency to confuse them with wishful thinking and delusions. Besides, through the approach they have removed or at least have lessened the paradoxes which make SD interesting. By doing so, they also make SD less thought-provoking. Second, by removing the intention element in SD, they lessen the responsibility of the self-deceiver in her SD which makes it difficult to assess its morality. Third, according to the intentionalists, the nonintentional accounts succumb to the problem of selectivity of SD. If

¹³ See Mark Johnston, "Self-deception and Nature of the Mind," in *Perspectives on Self-deception*, 64.

cases of SD were just following the whims of desires, impulses, and instinct, why is it that the self-deceiver decides on the circumstances of SD? For example, she chooses when and what to deceive herself about. The only answer is that she selects the object and the circumstances of SD. And this requires intention. We will see more of this in the next section.

3. Twisted Self-Deception within the Intentional Accounts

Very early in the debate, Demos has already pointed out cases of SD which are twisted. He has assumed that in terms of acquisition, the processes are just like those of the straight kind. When people lie to themselves, they can deceive themselves in favor of something pleasant or about something unpleasant. In both kinds of SD, there is a homogeneous explanation. Demos states:

My own analysis of self-deception follows a similar line. As with *akrasia*, there is an impulse favoring one belief at the expense of its contradictory; and the person who lies to himself, because of yielding to impulse, fails to notice or ignores what he knows to be the case.¹⁴

The long lists of intentionalists who came after Demos seem busier in explaining how SD is possible despite the paradoxes. It has led them to pay less attention to cases of twisted SD. Majority of them believe that, if there is deceptive intention, cases of SD can be explained homogenously. Nelkin echoes this assumption: "Intentionalists have a ready analysis of what is common to both straight and twisted cases: the self-deceiver forms the intention to deceive herself, succeeds, and the result is self-deception."¹⁵ Even though the SD belief that *p* is undesirable, the self-deceiver can still believe it because of her intention to deceive. Another homogenous explanation for all cases of SD is through its selectivity. Talbott's and Bermudez's respective accounts of SD are perfect examples of such a unified approach to SD.¹⁶ They argue that SD is selective. The self-deceiver chooses the circumstances when it is most appropriate to deceive herself. In fact, it would be disadvantageous for a creature were she to deceive herself only based on impulse or whims. If SD were nonintentional, humans would not survive, for desire would only

¹⁴ Demos, "Lying to Oneself," 594.

¹⁵ Dana Nelkin, "Responsibility and Self-Deception: A Framework," in *Humana Mentis Journal of Philosophical Studies* 20 (2012), 387.

¹⁶ See William, J. Talbott, "Intentional Self-Deception in a Single Coherent Self," in *Philosophy and Phenomenological Research* 55:1 (1995); and Bermúdez, "Self-Deception, Intentions, and Contradictory Beliefs," 309–319.

be after hedonistic goals. Indeed, they can choose to believe that which is unpleasant because they can intend to bias their belief or to desire to believe what is undesirable.

Both deceptive intention and the selectivity of SD arguments for twisted cases sound appealing. But desiring the undesirable is something hard to reconcile for it implies contradiction. Pears has already pointed this out when he talked about emotional cases of SD which are characteristically twisted. He has considered that, like the straight cases, they also have desirable goals. But while the latter's goal is geared towards the acquisition of a favorable belief, the former's acquisition of the unfavorable belief is just a means toward an ulterior goal:

So far, the assumption has been that in self-deception the motivation is always provided by a wish for some desirable goal. But is there always a desirable goal? And is there always a wish for it or are we sometimes merely programmed to go for it?

Consider self-deception caused by fear or jealousy. These emotions often lead people to form intrinsically unpleasant beliefs against the promptings of reason... In the case of fear, we may conjecture that the ulterior goal is avoiding the danger, and that it is best achieved by exaggerating it and so making quite sure of taking the necessary steps. Similarly, we may say that that the exaggerated speculations of jealousy, which are intrinsically unpleasant, the best way of making sure of elimination all rivals. In both cases the belief is a kind of bitter medicine.¹⁷

Even though there is a presumptive desirable goal of eliminating all rivals, desiring the undesirable belief that his wife is unfaithful is still problematic. Pears, in continuation, has spelled out the problem: "But neither fear nor jealousy cause people to want..., to form exaggerated beliefs. What, then, is the justification for postulating a wish in these cases?"¹⁸ In short, it is unthinkable for *S* to want the unwanted beliefs that jealousy and fear triggered. Since they are not wishful, it is simply difficult to see how emotion can trigger the desire to form disagreeable beliefs. His way out of this problem is to resort to the adaptive character of emotions:

¹⁷ Pears, *Motivated Irrationality*, 42–43.

¹⁸ *Ibid.*, 43.

There is presumably, a wish for the ulterior goal, safety or elimination of rival, but nature takes over at this point and sets up an emotional programme that ensures its achievements. The plan is nature and not the person's, and that is why the formation of the intrinsically unpleasant belief is not felt to be the object of the wish.¹⁹

Our emotions are adapted to respond in a manner appropriate to the stimuli. In the case above, jealousy might have exaggerated *p* as to compel *S* to embrace it in view of another goal. If it were the case, the acquisition of twisted SD differs greatly from the straight sort.

The rest of the intentionalists can still insist that deceptive intention can bring *S* to hold the undesirable belief that *p*, but they need to explain how it is possible. It is contradictory to desire to believe something undesirable. And Pears was right to invoke the role of emotion in those cases. But then, again, if it were the case, there would be more than one way of explaining the phenomenon.

Fitting twisted cases within desire-based explanation is problematic. It has even led people to ask whether they are really cases of SD, or whether they are special kinds of SD, or whether SD is really motivated. Ultimately, the problem of twisted SD involves the problem of accounting for the nature of SD. These questions, I think, can be sidestepped if we can find the proper place for emotions in SD which, as we have seen, have not been given attention by the intentionalists until Pears's discussion of the emotional cases.

4. The Nonintentional Twisted Self-Deception Accounts

Before dealing with motivating roles of emotions on SD, I will first review the major nonintentional approaches to twisted SD. Three major accounts can be identified from the literature: 1) the anxiety reliever account, 2) motivated biasing account, 3) and the purely emotional account.

The anxiety reliever account has been originally developed by Johnston and later modified by Barnes. Johnston has proposed that SD belief is generated by "*S*'s desire that *p* and his anxiety that not-*p*."²⁰ Barnes has found this inappropriate for twisted cases. The husband's SD belief that *p* (she is unfaithful) cannot be due to his anxiety that not-*p* (she is not unfaithful) because no anxiety would be reduced in such a case. Hence, to fit twisted cases, Barnes reformulated Johnston's as: "desire that *p* and anxiety that *q*" where *q* could refer to other worries. In the case of the jealous husband, *q* is

¹⁹ *Ibid.*, 44.

²⁰ See Johnston, "Self-deception and Nature of the Mind," 50–86.

the belief that “an esteemed colleague has a higher regard for her than for himself.”²¹ To reduce his anxiety that q , he ends up believing that his wife is unfaithful.

My objection to this account is that this might not be applicable with all cases of twisted SD. It may work in the case of the jealous husband, because the anxiety that q is greater than the anxiety that p . But it is difficult to use the account with other cases where it is hard to look for more anxious belief that q which can justify an anxious belief that p . Take the case of Trisha, the anorexic who holds the anxious belief that p , i.e., she is fat. It is simply hard to find a more anxious belief that q to justify that p . But since given that SD is an irrationality, the self-deceiver’s going for the less desirable belief seems conceivable. What may seem trivial to us, may not be to the twisted self-deceiver. Another objection is that of Scott-Kakures who argues against Barnes’s approach because rather than reducing the anxiety oftentimes, the preferred undesirable beliefs cause more anxiety.²² Mele has also raised his concerns because it is questionable whether all cases of SD involve anxious desire. For him, a self-deceiver can deceive herself even without being anxious about what she believes.²³

The second nonintentional approach is the one proposed by Mele and largely shared by Scott-Kakures. To explain SD, they have subscribed to the lay-hypothesis testing model proposed by Trope and Liberman,²⁴ Kunda,²⁵ Friedrich,²⁶ and Lewicka,²⁷ among others. The theory is based on the mechanism of the confirmation bias or what Baron calls “my-side” bias.²⁸ Confirmation bias (as a cognitive bias) functions independently of motivation. According to Kunda, people tend to confirm/favor their preexisting beliefs. The mere fact that a hypothesis is proposed or generated, people’s tendency is to conduct questions leading to its confirmation, which

²¹ Barnes, *Seeing Through Self-deception*, 36.

²² See Scott-Kakures, “Motivated Believing: Wishful and Unwelcome,” 368–369.

²³ See Mele, *Self-Deception Unmasked*, 55–56.

²⁴ Yaacov Trope and Akiva Liberman, “Social Hypothesis Testing: Cognitive and Motivational Mechanisms,” in *Social Psychology: Handbook of Basic Principles*, ed. by E. T. Higgins and A. W. Kruglanski (New York: Guilford Press, 1996), 239–270.

²⁵ Ziva Kunda, *Social Cognition: Making Sense of People* (Cambridge, Mass.: MIT Press, 1999).

²⁶ James Friedrich, “Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena,” in *Psychological Review* 100:2 (1993), 298–319.

²⁷ Maria Lewicka, “Confirmation bias: Cognitive error or adaptive strategy of action control?” in *Personal Control in Action: Cognitive and Motivational Mechanisms*, ed. by M. Kofta, G. Weary, and G. Sedek, (New York and London: Plenum Press, 1998), 233–258.

²⁸ Jonathan Baron, *Thinking and Deciding* (Cambridge, England; New York: Cambridge University Press, 2000), 203.

Kunda calls positive-test-strategy.²⁹ But with motivation, this tendency is bolstered.³⁰ In cases of straight SD, the application of the theory is quite direct. For example, without being motivated, the emperor would easily acknowledge his nakedness, or that Laura could easily recognize the evidence of her cheating husband, or that Sid can easily realize that his affection for Mary is not being reciprocated. Being motivated, the emperor, Laura, and Sid believe only the opposite.

Twisted cases can also be explained through the same biasing mechanisms of motivated beliefs. The idea is that *S* tends to confirm even the undesirable motivated beliefs since desires and/or emotions make the evidence supporting such beliefs more apparent to *S*. Scott-Kakures and Mele both advocate the use of hypothesis testing theories of Friedrich³¹ and that of Trope and Liberman.³² They both agree that emotions have an important role in twisted SD but do not seem to agree as to what this role consists in. On the one hand, Scott-Kakures considers it as a kind of motivation along with desire. As such, they share the functions of motivation in the biasing processes which he divides into two: 1) motivation triggers the hypothesis, thus initiating the cognitive biasing processes; and 2) motivation continuously supports the biasing processes as that of what happens in a typical hypothesis tester.³³ On the other hand, Mele reduces the role of emotions to being constituents of desire. From the very start, desire is actively biasing the processes by boosting the cognitive biasing mechanism which is directed at avoiding costly errors. This in turn, leads to confirmation of the motivated hypothesis rather than its rejection.³⁴

The third approach to twisted cases is proposed by Lazar and Dalglish. They argue that the effects of emotions on belief formation is obvious in both cases of SD. On one hand, Lazar exploits the fact of how 'mood shifts' can result to different interpretation of events or cases. It means that depending on a person's emotional states, an instance can be interpreted in different ways. This is more evident in twisted cases. For example, in the case of our jealous husband: "...in the grip of intense jealousy, (he) sees 'incriminating' evidence wherever he turns. In the grip of jealousy or rage, every aspect of his wife's behavior seems suspicious, while her affectionate

²⁹ See Kunda, *Social Cognition: Making Sense of People*, 113.

³⁰ See Alfred Mele, "When are we self-deceived?," in *Humana Mente Journal of Philosophical Studies* 20 (2012), 7.

³¹ Friedrich, "Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena," 298–319.

³² Trope and Liberman, "Social hypothesis testing: Cognitive and Motivational Mechanisms," 239–270.

³³ See Scott-Kakures, "Motivated Believing: Wishful and Unwelcome," 356–360.

³⁴ See Mele, *Self-Deception Unmasked*, 44–46.

behavior and consistent support are not given their due weight.”³⁵ On the other hand, Dalgleish proposes for emotionally biasing processes like that of Mele’s motivationally biased belief acquisition processes. Dalgleish explains that “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.”³⁶ It can only be surmised that since twisted cases are highly emotional, such biasing processes triggered by emotions are more appropriate for those cases.

There is such a possibility that emotions have an independent role in the biasing processes. However, the recent literatures reviewed by Bower and Forgas³⁷ regarding the interaction between emotion and cognition cannot support Lazar’s and Dalgleish’s claim for a sort of emotional roles (independent of desire) in the priming of the psychological mechanism that results in an acquisition of biased false beliefs. The lack of empirical evidence, however, might suggest that they have the same effect as desire in the triggering of the hypothesis that p which leads in the belief that p .

5. Emotions as Motivational Triggers of Hypothesis

There are other authors aside from Dalgleish and Lazar who have stressed the importance of the role of emotions in SD. A pioneer in this field would be De Sousa.³⁸ But his main interest has been to unravel why we often deceive ourselves about how we feel. He claims that emotions are intrinsically deceptive. This in turn may have an influence on our SD about our beliefs. In this sense, emotions have the same role as desire in motivating the self-deceiver into acquisition of her belief. He has not tackled, however, in what way they can motivate. And so, although he has not elaborated on the economy of SD belief acquisition, De Sousa assumes that emotions have a motivating influence in “self-deception focusing on belief.”³⁹ Other important theorists on emotional role in SD would be Sahdra and Thagard who approach SD through a computational model of emotional coherence. According to them, every judgment regarding a belief implies an emotional assessment or valence. One is self-deceived when the valence about a belief

³⁵ Lazar, “Deceiving Oneself or Self-Deceived? On the Formation of Beliefs Under the Influence,” 281.

³⁶ Dalgleish, “Once More with Feeling: The Role of Emotion in Self-Deception,” 110.

³⁷ Gordon Bower and Joseph Forgas, “Affect, Memory and Social Cognition,” in *Feeling and Thinking: The Role of Affect in Social Cognition*, ed. by Joseph Forgas (Cambridge: Cambridge University Press, 2010), 87–157.

³⁸ Ronald de Sousa, “Emotion and self-deception,” in *Perspectives on Self-deception*, 324–343.

³⁹ *Ibid.*, 327.

coheres with her subjective goals at the mercy of what should be the case or of the facts.⁴⁰

These studies on emotional roles on SD and the likes of them are relevant but they are not the roles of emotion that I am looking for in this study. My concern is much more basic than theirs in that I am dealing with the motivating roles of emotions in the biasing processes leading to the acquisition of SD belief which De Sousa assumed and that may be reason why Sahdra and Thagard see a valence in the belief of the self-deceiver.

Just like what De Sousa assumes, emotions have “a causal or motivational role” in SD.⁴¹ But just how emotions can motivate the acquisition of SD belief is still a matter of debate. Pears, as discussed above, differentiates between straight (wishful believing) types and twisted (emotional) types of SD which in dealing with the emotional types, the role of emotions is explained away by resorting to the adaptive character of emotions. The approach is a shortcut that leaves a lot of explanatory loopholes. Besides, he makes it appear that wishful types do not involve any role for emotions. Barnes’s account of SD, for her part, focuses more on the relief from anxiety (considering it as an emotion) as a motivating factor which accommodates both straight and twisted cases. But, as discussed above, anxiety may not be applicable to all cases of SD. Besides, if indeed the goal of the self-deceiver is to be relieved of her anxiety about a certain belief, in the twisted cases such belief is often a cause of greater anxiety. In the case of the jealous husband, we can question why he would prefer to believe that his wife is unfaithful to be relieved of the anxiety that the colleague has a higher regard for his wife than for him. Even if it is possible, it could be shown that the self-deceiver fell into that kind of irrationality because of some motivating influences on SD belief acquisition.

The accounts that represent a motivational approach fitting all cases of SD are those of Scott-Kakures and Mele whose model is that of lay-hypothesis testing based from Friedrich and Trope and Libermann.⁴² As noted above, Scott-Kakures has not differentiated between emotions and desire: they are both motivating influences responsible for SD belief acquisition. As such, they have dual functions of (1) triggering (the hypothesis) and (2) sustaining the processes of confirmation (of the hypothesis) leading to a biased acquisition of the belief represented by the hypothesis. Mele shares the same explanations in the acquisition of the biased belief that *p*. However, he focused more on the second aspect of motivational

⁴⁰ Baljinder Sahdra and Paul Thagard, “Self-Deception and Emotional Coherence,” in *Minds and Machines* 13 (2003), 213–231.

⁴¹ Ronald de Sousa, “Self-deceptive Emotions,” in *Journal of Philosophy* 75 (1978), 684.

⁴² Mele refers to this model as FTL theory as it is based on the theories of Friedrich, Trope, and Liberman.

functions. And he hardly speaks of the triggering of the hypothesis phase. Unlike Scott-Kakures who considers desires and emotions as motivation, Mele's understanding of motivation is solely equated with desire. He accepts that emotions can have a role but only as constituents of motivation. Apart from that, he doubts that emotion has an influence in the acquisition of SD belief parallel to that of desire.⁴³ In any case, both Scott-Kakures and Mele agree that in the second phase of hypothesis testing, the basic desire that motivates the hypothesis tester (in our case, the self-deceiver) is the avoidance of costly error. Mele argues that it is mostly a scheme of this unconscious desire which brings about the acquisition of SD belief. Scott-Kakures, however, maintains that the motivating influence of emotion and desire are continuously supporting the processes in support of that basic desire for avoidance of costly errors.⁴⁴

For Mele and Scott-Kakures, the role of emotion in motivationally biased belief acquisition is eclipsed by the role of desire. On the one hand, Mele reduces the role of emotions to being constituent of desire. A specific role for emotion in this sense is out of the picture. It can be seen on how he rejects Dalglish's assumption that emotion has an influence in the acquisition of SD belief parallel to that of desire.⁴⁵ On the other, Scott-Kakures's discussion is generic in the sense that emotions share this role with desire. And so, no specific role for emotion is elaborated.

In straight cases, whether emotion is involved seems irrelevant. Desire that *p* is sufficient to explain them: the self-deceiver wants *p* and so believes that *p*. In the twisted cases, the role of desire conflicts with the fact that it is hard to desire the undesirable. Here, the nonintentional motivationally biased belief accounts of SD based on lay-hypothesis testing theory accommodate well the twisted cases. If the account is right, once emotions trigger the hypothesis that *p* (e.g., "whether the wife is unfaithful," or "whether I left the burner flame on," or "whether I am fat"), testing for its confirmation is initiated. At least, in this triggering function, the role of emotion is obvious. It is jealousy that triggers the hypothesis of the wife's infidelity; fear that triggers the hypothesis that I have left the gas on or that there is a monster under my bed; and anxiety that I am fat. It is hard to see desire triggering such hypotheses.

In short, typical cases of twisted SD involved emotions influencing *S* to acquire belief that *p*. Here is a basic sketch of the process: emotions trigger a hypothesis that *p*; once *p* is triggered, it is proposed for confirmation. As *S* is biased towards *p* she ends up believing that *p*. We can see the case of the jealous husband fitting this description. In the sudden burst of jealousy, the

⁴³ Mele, "Emotion and Desire in Self-Deception," 174.

⁴⁴ Scott-Kakures, "Motivated Believing: Wishful and Unwelcome," 365

⁴⁵ Mele, "Emotion and Desire in Self-Deception," 175.

possibility of infidelity of his wife looms in his mind. He might start entertaining thoughts that could heighten his imagination of his wife's infidelity. He starts to look for grounds to support this hypothesis, ignoring contrary evidence. He ends up self-deceived that his wife is unfaithful. The same thing could happen in cases of fear. The attack of fear could spur the imagination to create a vision of a monster or a ghost. Once this is formed, a hypothesis that there is a monster, or a ghost could be formulated. The end-product is a SD belief that there is a ghost or a monster.

If there is a difference in the acquisition of the two kinds of SD belief, it pertains to how the hypothesis that p is triggered or generated. As emotions may trigger hypotheses whose objects are undesirable, so do desires (hunger, wants, hopes, lusts, etc.) most likely trigger pleasant hypotheses which can initiate the motivationally biased testing for confirmation. More often, associated pleasant emotions may also accompany such desires that can enhance the sustenance of testing for the confirmation of the hypothesis.

Given that, as opposed to Dalgleish and Lazar, I do not propose a different way of SD belief acquisition for twisted cases. In this account, emotions and desires are both motivating influences whose main function is to trigger a negative or a positive hypothesis, respectively. And so, the worry that there is not a homogenous approach to SD is somehow answered here. Emotions are a sort of motivating influence just like desires are. This account also complements Mele's FTL (Friedrich-Trope-Liberman) theory of lay-hypothesis testing. In explaining the FTL model, he has focused more in the second phase of theory proposed by the author, whereby an independent emotional role as suggested by Dalgleish and Lazar is denied. I must agree with him that in this second stage emotion is subsumed under desire in influencing the self-deceiver in his confirmatory quest of the triggered hypothesis. As with Scott-Kakures the function of motivation in the second phase is a sustenance of the hypothesis testing whose main adaptive ingredient is itself a desire to minimize or avoid costly errors. In all, the most specific role that we can ascribe to emotion is that of a motivating influence in the triggering or generating of the hypothesis that leads mostly to the acquisition of the twisted SD belief.

Conclusion

In this work, I have attempted to sketch a nonintentional account of SD that aims to address the problem regarding its twisted cases. I have argued that emotions are the main motivating influence in the acquisition of such SD beliefs. Specifically, their role lies in the triggering of the unfavorable hypothesis that p leading to the acquisition of the belief that p . Even though I posit emotions as triggers distinct from that of desires, I still maintain that the

process of SD belief acquisition is homogeneous, given that desires and emotions are both motivating influence in the triggering or generation of the hypothesis.

*Philosophy and Religious Programme, Faculty of Arts and Humanities,
University of Macau, Macau, SAR, China*

References

- Andersen, Hans Christian, "The Emperor's New Suit," in *Fairy tales of Hans Christian Andersen* (Auckland: Floating Press, 2014).
- Baghrmian, Maria and Anna Nicholson, "The Puzzle of Self-Deception," in *Philosophy Compass* 8:11 (2013).
- Barnes, Annette, *Seeing Through Self-deception* (United Kingdom: Cambridge University Press, 1997).
- Baron, Jonathan, *Thinking and Deciding* (Cambridge, England; New York: Cambridge University Press, 2000).
- Bower and Joseph Forgas, "Affect, Memory and Social Cognition," in *Feeling and Thinking: The Role of Affect in Social Cognition*, ed. by Joseph Forgas (Cambridge: Cambridge University Press, 2010).
- Bermúdez, José Luis, "Self-Deception, Intentions, and Contradictory Beliefs," in *Analysis* 60:4 (2000).
- Bok, Sisella, "The self deceived," *Social Science Information* 19:6 (1980), 931
- Dalglish, Tim, "Once More with Feeling: The Role of Emotion in Self-Deception," in *Behavioral and Brain Sciences* 20 (1997).
- Davidson, Donald, "Deception and Division," in *The Multiple Self*, ed. by Jon Elster (Cambridge: Cambridge University Press, 1985).
- De Sousa, Ronald, "Emotion and self-deception," in *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988).
- _____, "Self-deceptive Emotions," in *Journal of Philosophy* 75 (1978).
- Demos, Raphael, "Lying to Oneself," in *The Journal of Philosophy* 57:18 (1960).
- Friedrich, James, "Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena," in *Psychological Review* 100:2 (1993)
- Gardiner, Patrick, "Error, Faith, and Self-Deception," in *Proceedings of the Aristotelian Society* 70 (1970).
- Johnston, "Self-deception and Nature of the Mind," in *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988).
- King-Farlow, "Self-deceivers and Sartrean Seducers," in *Analysis* 23 (1963), 131–136; and Amelie Rorty, "The Deceptive Self: Liars, Layers, and

120 MOTIVATING INFLUENCE, EMOTION, SELF-DECEPTION

- Lairs," in *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988).
- Kunda, Ziva, *Social Cognition: Making Sense of People* (Cambridge, Mass.: MIT Press, 1999).
- Lazar, Ariela "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs Under the Influence," in *Mind* 108:430 (1999).
- Lewicka, Maria, "Confirmation bias: Cognitive error or adaptive strategy of action control?" in *Personal Control in Action: Cognitive and Motivational Mechanisms*, ed. by M. Kofta, G. Weary, and G. Sedek, (New York and London: Plenum Press, 1998).
- McLaughlin, Brian P., "Exploring the Possibility of Self-Deception in Belief," in *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988).
- Mele, Alfred, "Emotion and Desire in Self-Deception," in *Royal Institute of Philosophy Supplement*, ed., by Anthony Hatzimoysis (Cambridge: Cambridge University Press, 2003).
- _____, *Self-deception Unmasked* (United States: Princeton University Press, 2001).
- _____, "When are we self-deceived?," in *Humana Mente Journal of Philosophical Studies* 20 (2012).
- Nelkin, Dana, "Responsibility and Self-Deception: A Framework," in *Humana Mente Journal of Philosophical Studies* 20 (2012).
- Pears, David, *Motivated Irrationality* (New York: Oxford University Press, 1984).
- _____, "The Goals and Strategies of Self-Deception," in *The Multiple Self*, ed. by Jon Elster (Cambridge: Cambridge University Press, 1985).
- Sahdra, Baljinder and Paul Thagard, "Self-Deception and Emotional Coherence," in *Minds and Machines* 13 (2003).
- Scott-Kakures, Dion, "Motivated Believing: Wishful and Unwelcome," in *Nous*, 34:3 (2000).
- Talbott, William J., "Intentional Self-Deception in a Single Coherent Self," in *Philosophy and Phenomenological Research* 55:1 (1995).
- Trope, Yaacov and Akiva Liberman, "Social Hypothesis Testing: Cognitive and Motivational Mechanisms," in *Social Psychology: Handbook of Basic Principles*, ed. by E. T. Higgins and A. W. Kruglanski (New York: Guilford Press, 1996).