

Self-Deception in Belief Acquisition

Mario R. Echano

Abstract: Attempts to analyze self-deception (“SD,” from here on) have produced a series of articles and books trying to characterize the phenomenon and to resolve the problems involved in it. None has been found to satisfy the inquisitive minds as no analysis has been able to embrace all cases of SD. Alfred Mele is one of the leading scholars on SD. In the course of his works from 1982 to 2012, he has offered different formulations of SD, all aiming to accommodate all its instances. In this work, I examine Mele’s latest version of analysis of SD. I argue that his formulations exclude cases of twisted SD. Moreover, by his appeal to PEDMIN (Primary Error Detection and Minimization) that is involved in Lay Hypothesis Testing theory (LHT), he is courting a contradiction. Before delving into his set of conditions, I first situate the analysis in the problems that are involved in SD with desiderata {1}. In section {2}, I present the problem with Mele’s formulations of the analysis. In section {3} I dwell into why SD is acquisition of false belief. Section {4} is my justification of additional condition to Mele’s set of conditions. From sections {5} to {8}, I explain the relevance of the added condition to Mele’s existing set.

Keywords: Cognitive Bias, Motivated Believing, Self-Deception, Lay-Hypothesis Testing

Introduction

Attempts to analyze self-deception (“SD,” from here on) have produced a series of articles and books trying to characterize the phenomenon and to resolve the problems involved in it. None has been found to satisfy the inquisitive minds as no analysis has been able to embrace all cases of SD. Alfred Mele is one of the leading scholars on SD. In the course of his works from 1982 to 2012, he has offered different formulations of SD, all aiming to accommodate all its instances. In this work, I examine Mele’s latest version of analysis of SD. I argue that his formulations exclude cases of twisted SD. Moreover, by his appeal to PEDMIN (Primary

Error Detection and Minimization) that is involved in Lay Hypothesis Testing theory (LHT), he is courting a contradiction. Before delving into his set of conditions, I first situate the analysis in the problems that are involved in SD with desiderata {1}. In section {2}, I present the problem with Mele's formulations of the analysis. In section {3} I dwell into why SD is acquisition of false belief. Section {4} is my justification of additional condition to Mele's set of conditions. From sections {5} to {8}, I explain the relevance of the added condition to Mele's existing set.

{1} Desiderata

Plato seemed to have equated SD to 'lying to oneself.' In his *Cratylus*, Socrates speaks of self-deception as the worst kind of deception since the deceiver and the victim are one and the same: "... there is nothing worse than self-deception—when the deceiver is always at home and always with you..."¹ Indeed, 'lying to oneself' is the traditional ascription to SD. But equating it to SD implies contradiction or, at least, paradoxes.² Because lying is intentional, it seems impossible to intend to deceive oneself.³ In the recent

¹ Plato, *Cratylus*, trans. by Benjamin Jowett, in *The Collected Dialogues of Plato*, ed. by Edith Hamilton and Huntington Cairns (Princeton, New Jersey: Princeton University Press, 1961), 428d.

² Two paradoxes are involved in such an SD, namely: *dynamic* and *static*. The former consists in the difficulty of imagining how the self-deceiver can succeed in deceiving herself when she already knows what she is up to, and the latter is about the psychologically questionable state of the subject's holding of two contradictory beliefs. See Mario R. Echanó, "The Motivating Influences on Self-deception," in *Kritike: Online Journal of Philosophy*, 11:2 (December 2017), 104-120. See also Alfred Mele, "Real Self-Deception," in *Behavioral and Brain Sciences*, 20:1 (1997), 91-102.

³ To see the problem with equating 'lying to oneself' to SD in its plainest form, consider the case of my scheme of lying to a friend, Antonio. When I lie to Antonio that tomorrow is my birthday, my deceptive intention can be fulfilled only if it remains unknown to Antonio. He must not know that I intend to deceive him. If he does realize that I have such a plan, then I will not succeed on my attempt to lie. To push through my project of lying, I should be careful then not to give away my deceptive intention to Antonio. Moreover, if ever I succeed in convincing him of my lie, I now believe that tomorrow is not my birthday while he believes that tomorrow is my birthday. Once this scenario is applied to lying to oneself, the puzzles follow through. As in the case of lying just mentioned above: first, if I am going to lie to myself that it is my birthday tomorrow, I must not let myself know that I intend to deceive myself. Otherwise, once I discovered that I intend to deceive myself, I won't succeed. It seems to be an impossible feat for it is hard to keep from myself the knowledge of my deceptive intention, and the knowledge that what I have to convince myself about is false. Second, if ever I succeed in lying, I now believe that it is not my birthday tomorrow while at the same time I also believe that it is my birthday tomorrow. And this is simply a contradiction. Holding two contradictory beliefs at the same time, i.e. that p and that not- p (where ' p ' is the proposition expressed by 'it is my birthday tomorrow') seems intuitively impossible.

decades, the debate on SD centers on whether SD is 'lying to oneself' and consequently, on whether it involves deceptive intention.

I take the side of those who claim that most cases of SD do not involve deceptive intention. I do not exclude that 'lying to oneself' is a case of SD. Only that if indeed, it is possible, it could only be one of SD's various kinds.⁴ Besides, most instances of SD are not lying to oneself.⁵ For instance, we can accuse a jealous husband as self-deceived for being convinced that his wife is cheating (despite having no evidence to back up that belief), while professing that he does not want to believe that his wife is unfaithful.⁶

The consideration that most cases of SD are non-intentional is the best way to avoid the paradoxes. Since SD is not limited to lying to oneself, it is not necessarily intentional; and that it is not also necessary that a self-deceiver both believes p and not- p .⁷ With this, the conditions to qualify as SD would be less demanding. For SD to occur, it is sufficient that (1) one is motivated, and (2) that she believes something that is false.⁸

Setting aside that debate, another problem has cropped up in the theorists' characterization of SD, especially that both parties agree that (1). It is a problem when we see that there are cases of SD where, despite p being undesirable, the self-deceiver still believes it. Philosophers called them

⁴ Cited examples of lying to oneself are usually from popular literature, such as Aesop's *The Fox and the Grapes*, and Anderson's *The Emperor's New Clothes*, Shakespeare's *Tragedy of Othello*, etc. Here I use Tolstoy's *Kreutzer Sonata*. Some authors argue, however, that these cases can also be explained as non-intentional.

⁵ On the one hand, 'lying' has generally been considered intentional. See, for example, the following: Augustine, "Lying," trans. by Sister Mary Sarah Muldowney, in *Treatises on Various Subjects*, ed. by Roy Deferrari, Vol. 16 of *The Fathers of the Church* (Washington, D.C.: Catholic University of America Press, 1952), 53-110; Raphael Demos, "Lying to Oneself," in *Journal of Philosophy*, 57:18 (1960), 588-595; Sisela Bok, *Lying: Moral Choice in Public and Private Life* (New York: Random House, 1978); J. E. Mahon, "The Definition of Lying and Deception," in *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), ed. by Edward N. Zalta, <<https://plato.stanford.edu/archives/win2016/entries/lying-definition/>>. For a contrary claim, see D. Fallis, "Are Bald-Faced Lies Deceptive After All?" *Ratio*, 28 (2015), 81-96. Most of these authors subscribed to the dictionary definition that 'to lie' is 'to make a false statement with the intention to deceive' (*Oxford English Dictionary*, 1989). On the other hand, deception is not necessarily intentional. OED, 1989 says of "to deceive" as to cause to believe what is false," what matters in deception (of others) is the falsity of belief. For a thorough review of the differences between lying and deception, see Mahon, "The Definition of Lying and Deception."

⁶ He does not intend to believe that his wife is cheating but believes it anyway. Vasyla Pozdnishef, the husband who killed his wife out of jealousy in Tolstoy's *Kreutzer Sonata* may be such a case.

⁷ Here " p " refers to the proposition and " S " refers to the subject.

⁸ This position is called deflationary SD. For further explanation on how deflationary SD avoids the paradoxes, see Echano, "The Motivating Influence of Emotions on Self-deception," 108-110. See also Alfred Mele, "Self-Deception and Delusion," *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation*, ed. by T. Bayne & J. Fernandez, (New York: Psychology Press, 2009), 55-56.

twisted cases to contrast them with the straight ones where one believes p because of the desire for p .⁹ The main question is: how can something undesirable be desired after all? The problem of twisted cases of SD puts into question that SD is motivated by desire.

I still maintain that SD is like most phenomena of motivated biased believing that are considered in social and cognitive psychology. By being motivated (also called 'hot') SD is distinguished from the cognitive ('cold') biased believing which are adaptive and unconscious. In this approach to SD, I am following the path trod by Talbott,¹⁰ Johnston,¹¹ Lazar,¹² Mele,¹³ Scott-Kakures,¹⁴ Bermúdez,¹⁵ Nelkin,¹⁶ and Dolcini,¹⁷ among others. The influence of motivation on cognitive biasing processes is crucial in these philosophers' approaches to SD.

One application of motivated biasing process that is parallel to belief acquisition in SD is that of Lay Hypothesis Testing (LHT) theory.¹⁸ LHT theory claims that people tend to confirm their favored hypothesis by the mere fact of it being proposed as a hypothesis.¹⁹ This is thus an application of the confirmation bias.²⁰ It assumes that before S ends up believing p , she must

⁹ See Echano, "The Motivating Influence of Emotion on Self-deception, 104-117, for the characterization of such cases of SD.

¹⁰ See William, J. Talbott, "Intentional Self-Deception in a Single Coherent Self," in *Philosophy and Phenomenological Research*, 55:1 (1995); and José Luis Bermúdez, "Self-Deception, Intentions, and Contradictory Beliefs," in *Analysis*, 60:4 (2000), 309-319.

¹¹ See Mark 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), 63-91.

¹² See Ariela Lazar, "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs Under the Influence," in *Mind*, 108: 430 (1999), 265-290.

¹³ Alfred Mele, *Self-deception Unmasked* (United States: Princeton University Press, 2001).

¹⁴ See Dion Scott-Kakures, "Motivated believing: Wishful and Unwelcome," in *Nous*, 34:3 (2000), 348-375

¹⁵ Bermúdez, "Self-Deception, Intentions, and Contradictory Beliefs."

¹⁶ Dana Nelkin, "Responsibility and Self-Deception: A Framework," in *Humana Mentis Journal of Philosophical Studies*, 20 (2012), 384-406.

¹⁷ Nevia Dolcini, "The Pragmatics of Self-Deception," in *New Directions in Logic and the Philosophy of Science*, ed. by Laura Felline, Antonio Ledda, and Francesco Paoli (United Kingdom: College Publications, 2016), 67-76).

¹⁸ In contrast with scientific (statistical) hypothesis testing, LHT or everyday hypothesis testing in social psychology is the usual unconscious way of making up one's mind. See Ziva Kunda, *Social Cognition: Making Sense of People* (Cambridge Mass.: MIT Press, 1999), 109-121.

¹⁹ A recent version of this theory can be seen in A. P. Gregg, N. Mahadevan, and C. Sedikides, "The SPOT Effect: People Spontaneously Prefer Their Own Theories," in *The Quarterly Journal of Experimental Psychology*, 70-6 (2017), 996-1010. They claim that people tend to spontaneously favor their own hypothesis by mere association that it is theirs or that they can relate to it.

²⁰ For further discussion on this, see J. Klayman and Y. Ha, "Confirmation, Disconfirmation, and Information in Hypothesis Testing," *Psychological Review*, 94, (1987), 211-228; James Friedrich, "Primary error detection and minimization (PEDMIN) strategies in social

first posit it as a hypothesis. Once posed, it activates positive test strategies which are already biased towards its confirmation.²¹ My approach to SD is an application of LHT to explain that SD is non-intentional, and that it is motivated by desires and/or emotions.

Mele always maintains that in entering SD in acquiring a belief that *p*, desire for *p* is the main motivation, even in twisted cases. As stated above, this is problematic. I²² argued that both desire for *p* and/or emotion towards *p* are motivating influences in the acquisition of belief that *p*. This explains how a self-deceiver ends up believing that *p* despite its undesirability. In this work, I do not intend to present my own list of conditions for SD-belief acquisition. Rather, I will comment and introduce a modification on Mele's set of conditions for entering SD in acquiring a belief that *p*.²³ I argue that without this modification, his list of conditions for entering SD in acquiring a belief that *p* would still exclude the twisted cases.

Some words of caution before going through the coming sections. In this work, I limit myself with SD in belief acquisition. Being self-deceived in feelings, aspirations, dreams, desires, etc., is beyond its scope. Second, entering SD in acquiring or retaining a belief that *p* must be distinguished from perpetuating oneself in the state of SD. I limit this work in the consideration of entering SD in belief.²⁴ Third, of the two ways on entering SD, I further limit this work into examining the conditions for entering SD in belief acquisition.²⁵ The list of conditions that will be examined in this work pertains only to that of entering SD in acquiring a belief that *p*.

cognition: A reinterpretation of confirmation bias phenomena," in *Psychological Review*, 100:2 (1993), 298-319; and Gregg, Mahadevan, and Sedikides, "The SPOT Effect."

²¹ For example, once asked whether *p*, we asked that *p* (e.g., that the lecture is boring, we tend to verify whether it is boring, not whether it is not boring?) We are 'wired' to ask questions leading for confirmation.

²² See Echano, "The Motivating Influence of Emotion on Self-deception."

²³ Having written about twenty articles/papers and two books on the topic, Mele is one of the leading authorities on SD.

²⁴ I am following Mele in these aspects. See Alfred Mele, "When are we self-deceived?" *Humana Mente Journal of Philosophical Studies*, 20 (2012), 1-16.

²⁵ There are two ways of entering SD: that of acquiring a belief that *p* and that of retaining a belief that *p*. Here is an illustration from Mele: "Sam has believed for many years that his wife, Sally, would never have an affair. In the past, his evidence for this belief was quite good. Sally obviously adored him; she never displayed a sexual interest in another man; . . . she condemned extramarital sexual activity; she was secure, and happy with her family life; and so on. However, things recently began to change significantly. Sally is now arriving home late from work on the average of two nights a week; she frequently finds excuses to leave the house alone after dinner; and Sam has been informed by a close friend that Sally has been seen in the company of a certain Mr. Jones at a theater and a local lounge. Nevertheless, Sam continues to believe that Sally would never have an affair. Unfortunately, he is wrong. Her relationship with Jones is by no means platonic." Mele, "Real Self-Deception," 95-96.

In this case, Sam did not acquire a new belief which is false. He acquired it as a true belief. Later, however, there was a change of situation as Sally is no longer faithful. And yet,

{2} Jointly Sufficient Conditions for Entering Self-Deception

According to Mele, *S* enters SD 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 non-deviant cause of *S*'s acquiring the belief that *p*
- (4) The body of data possessed by *S*, at the time, provides greater warrant for not-*p* than for *p*
- (5) *S* consciously believes, at the time, that there is a significant chance that not-*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.²⁶

The above list aims to provide jointly sufficient conditions for entering all cases of SD. Condition (1) states that SD is about the falsity of the belief that *p* which *S* acquired; conditions (2), and (3) pertain to the motivationally biased treatment of data regarding *p*; and conditions (4), (5) and (6) stress the characteristics of tension.

Since I claim that this list is not enough to address the problems raised by SD's twisted cases, I am suggesting a modification on the list. I add another condition to more conveniently accommodate the cases of twisted SD, and to provide stronger support for the intuition that SD is motivated. This added condition is formulated thus: *S*, triggered by desire or emotions, generates a hypothesis that *p*. And so, the modified list will look like this:

despite evidence to the contrary, he still wants to maintain the former belief. He is self-deceived in retaining the belief that Sally is faithful.

²⁶ Mele, "When are we self-deceived?"

- [1] The belief that p which S acquires is false
- [2] S , triggered by desire or emotions, generates a hypothesis that p
- [3] S treats data relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way
- [4] This biased treatment is a non-deviant cause of S 's acquiring the belief that p
- [5] The body of data possessed by S , at the time, provides greater warrant for not- p than for p
- [6] S consciously believes, at the time, that there is a significant chance that not- p
- [7] 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.²⁷

As could be observed above, I maintained Mele's list except for [2]. Indeed, it is understandable that his list would be more focused on setting up conditions for straight cases because twisted cases were not an issue in his earlier writings where he presented prior versions of the list.²⁸ Still, when the problems raised by the twisted cases cropped up, he insisted that they are also complying with the conditions. In trying to fit those cases within the conditions in the list, he resorted to the Primary Error Detection and Minimization theory (PEDMIN)²⁹ which is, for him, an essential aspect of LHT theory in showing that such cases are as motivated as the straight ones. This move, however, is problematic. The main reason is that if twisted cases merely rely on the desire to avoid the cost of being mistaken, as the PEDMIN

²⁷ To differentiate the modified list from the original, I use "[#]" sign instead of Mele's "(#)."

²⁸ From Alfred Mele's "Self-deception," in *Philosophical Quarterly*, 33 (October 1983), 366-377, to his "Approaching Self-Deception: How Robert Audi and I part company" in *Consciousness and Cognition*, 19:3 (2010), 745-750, the original four-point jointly sufficient conditions are presented. Amidst objections to this four-point conditions, he introduced the last two in Mele, "When are we self-deceived?"

²⁹ PEDMIN was first proposed by Friedrich, "Primary error detection and minimization (PEDMIN) strategies in social cognition." It explains why we tend to confirm a hypothesis. It claims that we have this adaptive tendency to avoid being mistaken about a hypothesis. People's tendency to confirm hypothesis is motivated by the basic desire to avoid costly error.

theory claims, then such explanation runs in conflict with his condition (3). There would only be a sort of deviance causality involved in such cases of SD as the desire to avoid the cost of error would be the main motivation in confirmatory testing of the hypothesis. If that is so, then twisted cases cannot be counted as SD. Furthermore, this desire in PEDMIN could also be a threat to his claim that all cases of SD (twisted and straight) are products of the desire for p . Cases of SD could all just be reduced to being a product of that basic desire in PEDMIN, and thus, would be violating the non-deviance condition.

To further see where Mele's list falls short, let us test some of the known twisted cases against his list of conditions. Let us look at the cases of Vasyla Pozdnishef, the jealous husband who murdered his wife in Tolstoy's *Kreutzer Sonata*,³⁰ and of Adonis's fear of ghosts, as examples. It would be incomprehensible to take the desire of those subjects to avoid the cost of being mistaken about their respective hypotheses that p (e.g. that the wife is unfaithful; or that there is a ghost) as the main motivations for their confirmation. In the first place, without being generated, how can those hypotheses come about? It would also appear that the desire in PEDMIN is disassociated with the desire for p . As such, the desires, of which jealousy or fear is a component, are no longer the causes of the self-deceiver's acquiring the belief that p . And so, those cases do not really conform to the non-deviance condition. There is another problem with Mele's reliance with PEDMIN; Haselton and Buss's Error Management Theory (EMT) claims that, at least, some cases of error minimization and management are but adaptive, and has nothing to do with any motivation to avoid error.³¹ It is premature to install PEDMIN as the main reason for S 's tendency to confirm the hypothesis that p . The introduction of the additional condition (i.e. my condition [2]), avoids this problem. Moreover, it assures that the process of self-deceptive belief acquisition is non-deviant.

Before dwelling with this modification further, I will first comment on the condition [1] in the next section, and then I will discuss the role of condition [2] in sections 4. To set the direction of the rest of this work, it is convenient to point out at this juncture that the process of acquisition of a self-deceptive belief begins with the added condition [2]: S , triggered by desire or emotions, generates a hypothesis that p . The process, eventually,

³⁰ Leo Tolstoy, *The Kreutzer Sonata* (Greenwich, CN: Fawcett Publications, 1960).

³¹ M. Haselton and D. Buss, "Error Management Theory: New Perspective on Biases in Cross-Sex Mind Reading, *Journal of Personality and Social Psychology*, 78:1 (2000), 522-523. Like PEDMIN, the Error Management Theory (EMT) can also be applied in LHT theory. However, it claims the opposite of PEDMIN. According to EMT, there are cases where there is no such basic desire to avoid costly error is involved in LHT. Hence, at least some cases of error minimization or management is merely adaptive and purely cognitive.

ends up with *S* believing that *p*. I will discuss the rest of the conditions in the remaining sections.

{3} The False Belief Condition

Condition [1] puts SD opposite to that of knowledge traditionally analyzed as justified true belief (JTB) in the sense that it indicates that SD is not concerned with knowledge acquisition. Although Dolcini disagrees with this condition, she rightly expresses SD's opposition to knowledge as JTB by saying that SD is a failure to acquire knowledge:

Indeed, C^{FB} (or false belief that *p* condition) is a condition about the truth value of *p*, so that it shifts the observer's attention from a doxastic level to an epistemological level of analysis: entering self-deception looks like a process by which the subject fails to acquire knowledge, where knowledge is intended in the traditional sense of (at least) justified true belief.³²

Condition [1] also has nothing to do with the dynamics of self-deceptive belief acquisition.³³ It is independent of the processes leading to SD. And yet, it is the most basic of all jointly sufficient conditions. In one of Mele's earliest articles in SD, he posed this question: "What must be added to false belief in order to yield a condition of self-deception?"³⁴ This implies that even if *p* is acquired in a motivated fashion, if it happens that *p* is true, there is no SD.

There are at least three groups of theorists who disagree with this condition. The first group pertains to those who maintain that SD is an analogue of interpersonal deception, and thus, considers that condition (1) is not enough.³⁵ They require that the self-deceiver both believe that *p* and that not-*p*. The self-deceiver's possession of contradictory beliefs is a consequent of having interpersonal deception as model for SD. The second group

³² Dolcini, "The Pragmatics of Self-Deception," 6

³³ Nelkin's version of this condition is simply formulated as "*p* is false." Such formulation would be more preferable if one is concerned with the etiology of the self-deceptive belief acquisition. See Nelkin, "Responsibility and Self-Deception: A Framework."

³⁴ Mele, "Self-Deception," 371

³⁵ Donald Davidson, "Deception and Division," in *The Multiple Self*, ed. by Jon Elster (Cambridge: Cambridge University Press, 1985); and Francis Pears, "The Goals and Strategies of Self-Deception," in *The Multiple Self*, ed. by Jon Elster (Cambridge: Cambridge University Press, 1985); Amelie Rorty, "The Deceptive Self: Liars, Layers, and Lairs," in *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988); and Bermúdez, "Self-Deception, Intentions, and Contradictory Beliefs," in *Analysis*, 60:4 (2000).

pertains to the accounts of SD that treat p independent of its truth value. Talbott and Dolcini hold such an account.³⁶ For Talbott, what matters in SD is for the self-deceiver to be motivated to believe that p whether it is true or not. For Dolcini, one can still be considered self-deceived even if she is in possession of a true belief that p , provided she also meets the deviation condition (C_D), tension condition (C_T) and motivation condition (C_M). The third group is constituted by those who claim that the false belief condition is too much of a demand for SD. A few of them are Audi, Bach, Funkhauser, Gendler, and Rey.³⁷

In support of the condition [1], it would also be helpful to mention here a survey, conducted by Mele among his philosophy students in Florida State University, which aims to establish that SD, pre-theoretically understood, includes cases where the self-deceived holds a false belief. He used this survey against Audi's claim that p 's falsity is not required for SD. The survey provides an empirical support to his claim that self-deceived persons do hold the false belief that p .³⁸

In sum, condition [1] does not imply that the falsity of p has special importance for the processes of SD. It is independent of the motivationally or emotionally biasing dynamics. It immediately discounts as SD the improbable belief that p if it is arrived at because of S 's motivationally biased treatment of the data. Perhaps the subject might be self-deceived about some beliefs supporting the belief that p . We can recall the case of Laocoön; he merely got 'lucky' that his belief was true. All the elements of the sufficient conditions for entering SD were present except for the condition [1].

Just as in in the case of knowledge, however, it is not enough for a belief that p to be true to yield knowledge; it is also not enough for p to be false for one to yield SD. Whereas in the case of knowledge, the justification of true belief is required to eliminate lucky guesses that happen to be true as knowledge, other conditions must also be set up for false belief to be considered SD-belief. The rest of the conditions assures that inadvertent false belief that p is a kind of SD.

³⁶ See Talbott, "Intentional Self-Deception in a Single Coherent Self" and Dolcini, "The Pragmatics of Self-Deception."

³⁷ Robert Audi, "Self-deception, rationalization, and reasons for acting," *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988), 92-120; Kent Bach, "An Analysis of Self-Deception," in *Philosophy and Phenomenological Research*, 41 (March 1981), 351-37; E. Funkhauser, "Do the Self-Deceived Get What They Want?" *Pacific Philosophical Quarterly*, 86:3 (2005), 295-312.

³⁸ Mele, "Approaching Self-Deception: How Robert Audi and I part company."

{4} Generation of the Hypothesis that p Condition

Adding the condition [2] to Mele's list has at least three advantages. First, this condition specifies how desire and emotions exert a motivating influence on the acquisition of self-deceptive belief. While desires trigger desirable hypothesis, emotions are triggers for the undesirable ones. Condition [2] aims to show that straight cases are products of the desire for p . At the same time, it attempts to demonstrate that cases of twisted SD are products of the motivating influences of emotions. Such explanation is clearly distinct from Mele's account where he assumed that twisted cases can be explained through that basic desire to avoid the cost of being mistaken in LHT. Second, condition [2] is important in showing that the process of SD is motivated against the doubts on whether all the confirmatory tendencies of testing the hypothesis are motivated. In section 2, I mentioned that EMT purports that some cases of avoidance of costly errors in testing the hypothesis have nothing to do with desire to avoid the cost of being mistaken. Condition [2] assures that SD's being motivated does not altogether rely on the actual testing of the hypothesis but on its generation. Lastly, this condition assures that the process of self-deceptive belief acquisition is non-deviant. In the sub-sections that follow, I will attempt to elucidate on those first two reasons. And, I will discuss in subsequent section the last advantage where condition [2] offers a guarantee that the process is non-deviant.

{4.1} The Role of Desires and Emotions

Like Mele, I also take the LHT theory as a model for explaining how motivation becomes responsible for S ending up with false biased belief that p .³⁹ Unlike him, I prefer to stress, through condition [2], the respective hypotheses that either emotions or desires trigger. Below, I sketch briefly the arguments for the motivating influence of emotions in twisted SD, and of desire in the acquisition of straight self-deceptive belief.⁴⁰

The LHT theory, used by Mele⁴¹ and Scott-Kakures,⁴² implies the motivating influences of desire and emotions on the acquisition of biased (false) beliefs. The theory has been commonly used in social psychology; and it attempts to explain how people tend to confirm, rather than reject,

³⁹ See Mele, *Self-deception Unmasked*.

⁴⁰ In a recent paper, Lauria and Preissmann attempts to provide a unified account of SD through the role of emotions which is thus a more radical claim than mine. See Federico Lauria and Delphine Preissmann, "What Does Emotion Teach Us About Self-Deception: Affective Neuroscience in Support of Non-Intentionalism," in *Les Ateliers de L'Ethique*, 13-2 (2018), 70-94.

⁴¹ Mele, *Self-deception Unmasked*.

⁴² Scott-Kakures, "Motivated believing: Wishful and Unwelcome."

hypothesis once it is posed for testing, mostly through the use of the positive test strategy. According to Trope and Liberman, LHT theory has two parts, namely the generation of the hypothesis, and its testing.⁴³ This condition [2] refers to the first part, that is, the generation of a hypothesis that p following LHT theory. I argue that to enter SD in acquiring a belief that p , it is necessary that S first generates a hypothesis that p . Thus, 'whether p ?' is prior to the acquisition of the belief that p . Trope and Liberman suggest some instances on how hypotheses are generated:

Sometimes, others provide hypothesis by asking questions or making assertions that raise possibilities the individual is then motivated to test. For example, a direct question or an assertion regarding an acquaintance's friendliness may lead one to test the hypothesis that this person is actually friendly

In many real-life situations, however, hypotheses are spontaneously generated by the hypothesis tester. In deciding whether and how to pursue their goals people often want to know if necessary preconditions have been met. Therefore, people formulate hypothesis about these preconditions. For example, someone trying to hire a sales-person may try to determine whether an applicant is extroverted, believing that extroversion is important for success in sales.⁴⁴

Mele also mentions these two ways by which hypothesis generation is made possible: first, others can generate it for S by suggesting p ; and second, S may have generated p by herself.⁴⁵ Other than mentioning them, he seems to ignore the relevance of the generation of hypothesis in the process. My claim is that in cases of entering SD in acquiring a belief that p , S forms the proposition because of desire and/or emotions. Whether it is suggested by others or generated by S herself, what is important is that p has an appeal to emotion or desire so as to trigger the process that leads to self-deceptive acquisition of the belief that p .

Mele's list of the conditions lacks my proposed condition [2] because, although he makes use of the LHT theory, his primary goal is to explain how

⁴³ Y. Trope and A. Liberman, "Social Hypothesis Testing: Cognitive and Motivational Mechanisms," in *Social Psychology: Handbook of Basic Principle*, ed. by E. Higgins and A. Kruglanski (New York: Guilford Press, 1996), 239.

⁴⁴ *Ibid.*, 240.

⁴⁵ Cf. Mele, *Self-Deception Unmasked*, 33

motivation leads one to confirm, rather than reject, the favored hypothesis. It is his non-intentional alternative explanations to the intentionalists' accounts. At the same time, he is concerned with providing a unified explanation of the phenomenon of SD. He has been preoccupied with giving a homogenous explanation for all cases of SD through the desire that p . This is, in turn, the reason why he found it hard to locate a place for an independent role of emotions in twisted cases. Hence, even though he also considered several hypotheses regarding the role of emotions, he argued mostly for emotions' role as constituent of desire. Against this, I⁴⁶ have argued that while desires trigger desirable hypothesis, emotions are triggers for the undesirable ones. Condition [2] aims to point out that straight cases are products of the desire for p ; it also shows that twisted cases are products of emotions towards p .

As noted above, Mele's list of conditions is suited more for explaining straight cases of SD. It tries to accommodate twisted cases by a recourse to PEDMIN. Again, such approach makes the basic desire to avoid the cost of being mistaken to be the immediate cause of one's acquisition of self-deceptive belief. As a result, desire for p , as Mele purports, is not the cause of the acquisition of twisted self-deceptive belief.

If the initial phase of generation of the hypothesis that p condition has been considered seriously enough, twisted cases could be offered a better explanation other than being subsumed under that basic desire not to commit costly errors in PEDMIN. This additional condition [2] aims to better accommodate the cases of twisted SD. It specifies, from the very beginning, the motivating influences in the acquisition of self-deceptive belief, thereby, assuring that the process of self-deceptive belief acquisition is either motivated by desire or by emotions. The next sub-section discusses further the relevance of the condition [2] in Mele's list.

{4.2} The Role of Generation of Hypothesis that p in Self-Deception

Again, the generation of the hypothesis that p is the initial phase in the process that leads to the acquisition of the biased belief that p . The added condition [2] states that for p to settle on S 's mind, it must first come as a hypothesis.

In his earlier discussions of the problems of SD, Mele was not using the LHT theory,⁴⁷ thus, the belief that p which starts in being a hypothesis that p has not yet been formulated. But even when he began to use the LHT theory, he did not pay much attention to the generation of hypothesis phase of LHT in the process of self-deceptive belief acquisition. In contrast, Scott-Kakures

⁴⁶ Echano, "The Motivating Influence of Emotions in Self-Deception."

⁴⁷ From 1993 to 1997, Mele has not considered the LHT in his works on SD. It was only after 1997, that he considered the importance of that theory in SD.

emphasized this initial phase, when arguing that the generation of the hypothesis that p also determines S 's goals and values that direct hypothesis testing. The workings of this initial phase could be understood in this way: if the proposed belief that p has been motivationally or emotionally loaded from the very start, then the hypothesis tester's confirmatory tendency follows spontaneously. The subsequent job of motivation in the biased confirmatory testing of p follows on cue.

It is true that my additional condition to Mele's list is already implied in his formulation of the list. In fact, it is assumed in his conditions (1) and (2) because in order for S to treat the truth value of the belief that p and to acquire it, belief that p must come first as a hypothesis that p . As I have pointed out already, there is a need to put this explicitly. The added condition not only specifies the kind of motivating influence involved in SD, it also assured that there is a non-deviant causal connection between the motivating influences of desire and emotions and the acquired self-deceptive belief.

The importance of desire and emotion at this stage is crucial whether a proposition is suggested by others or generated by the self-deceiver herself. Following the model of LHT, desire and emotion trigger the generation of belief that p . Take again the case of the emperor in Anderson's *The Emperor's New Clothes*. His is an example of the first instance of hypothesis generation of p . The tale is a case of other-deception that involved SD. The emperor's desire for new and unique dress leads to his belief that he is indeed wearing one, when in fact he is naked. The swindlers tricked him into embracing the false belief that he is wearing a new and unique dress by appealing to his desire for a new dress. Though the swindlers are responsible for the deception, he shares the responsibility by overlooking the evidence for his nakedness in favor of his desire for the dress. He is deceived by the swindler, and at the same time, he is self-deceived. He desired it even before it was suggested to him. Otherwise, he would not have fallen into believing that he is wearing a new suit if, for him, wearing the new suit was not desirable.

In brief, when p is generated by S or suggested to S , it has to come as a hypothesis. There may be evidence for and against p . But more often than not, there are more evidence for not- p rather than for p . As in the case of the emperor, there are evidence for his nakedness and only the testimony of the swindlers for his being wonderfully dressed. Or in the case of the Vasyla, there are more evidence for his wife's fidelity rather than for her infidelity. Still, in these cases, S has to weigh up the opposing evidence in order to finally believe or reject that p once their respective thresholds have been reached. And because of the influence of desire and/or emotions, she may easily reach the acceptance of p threshold. We say that she immediately jumps to conclusion without properly reviewing the evidences whereby in the absence of these motivating influences, she could easily realize that p is false.

Whichever way it is, she still has tested the hypothesis that not- p albeit with a bias towards its rejection (condition [3]).

The hypothesis that p may be about something unpleasant would still be confirmed not only due to this tendency to avoid being mistaken about the hypothesis that p , but because it is, in the first place, triggered by emotions. And so, with this added condition [2], even if the evaluation of the hypothesis in LHT turns out to be adaptive and mere cognitive, as Haselton and Buss proposed in their EMT, the motivated character of SD would be assured since the generation of the hypothesis has been motivationally charged to sustain its subsequent evaluation.⁴⁸

Sometimes, subjects generate their self-deceptive belief out of desire/emotion, and independently of others' suggestion. We can conjecture that, in typical cases of straight SD, S 's desire for an object would spur him to form p . Thus, desire generates p -hypothesis. Once generated, it is no longer a mystery that he would desire that p since it was formed by desire in the first place. Such may be the example of the emperor whose desire for a new suit leads him to form a hypothesis that he has a wonderful suit on despite the utter evidence of his nakedness. Again, once the motivated hypothesis is running, the tendency would be to confirm it. Same could be said of the jealous husband, Vasyla. His jealousy triggers the hypothesis that his wife is unfaithful. Then, the tendency to confirm it was in the bag, not only because of the inherent motivation to confirm the hypothesis but also, because of the jealousy which prevents him from being objective or rational and could also be boosting the adaptive tendency to favor the motivationally or emotionally charged hypothesis. In the case of fear, as that of Adonis, a child who believes that there is a monster lurking under his bed, the attack of fear could spur the imagination to create a vision of a monster or a ghost. Because of that, the hypothesis that there is a monster or a ghost could be formed. The end-product is a self-deceptive belief that there is a ghost or a monster. It would be hard to explain that the belief on a monster under his bed has been triggered by desire. In all these, condition [2] renders cases of twisted SD fitter within the motivational explanation.

{5} Treatment of Motivationally and Emotionally Generated Hypothesis

The condition [3] is what follows after the generation of the hypothesis that p . In LHT, it pertains to the actual testing of the hypothesis. When the emotionally or motivationally charged belief that p is suggested to

⁴⁸ M. Haselton and D. Buss, "Error Management Theory and the Evolution of Misbeliefs," *Behavioral and Brain Sciences*, 32:6 (2009), 110-111.

S, the expected outcome will be a biased treatment of data relevant, or at least, seemingly relevant to the data favoring the confirmation of the hypothesis. Following the processes of the LHT theory as explained above, condition [3] is mainly a work of motivation. Above in {4}, we have seen how the LHT theory helped in showing the processes in the acquisition and retention of the false belief that *p*. There, I discussed how the motivated and non-motivated processes lead to the acquisition (or retention) of the false belief that *p*. I noted that the confirmation bias which underlies the theory of LHT is usually unmotivated. The acquired biased false belief, resulting from cold biases and heuristics, such as confirmation bias, does not qualify for SD precisely for lack of the motivating influence. However, if in the first place, the proposition is already motivationally or emotionally charged, the treatment of data relevant to the proposition will be motivationally and emotionally charged as well.⁴⁹ Both EMT and PEDMIN theories would also be affected by such prior motivating influences. Tendencies to confirm a hypothesis would be boosted because of those motivating influences.

Condition [3] points to the fact that once we are motivated by desire or emotions, we usually end up believing something that is motivationally charged. It makes us vulnerable to false belief acquisition. Desire and emotions color how we treat the data relevant to its truth value. And since most cases of twisted SD involved beliefs which are emotionally charged when they are proposed to *S*, they are also strong candidates for confirmation. The task of confirming even those unpleasant propositions would be facilitated by those motivating influences that boost the sources of confirmation bias. This condition, like the next one, heavily relies on the added condition [2]. This additional condition does not only direct the successive confirmatory testing of the hypothesis, it also assures that the process is non-deviant.

{6} The Non-Deviance Condition

Condition [4] refers to the non-deviance of the causal connection in the biasing process and the acquisition of belief that *p*. To count as SD, the biasing process which starts from condition [2], i.e., the cause of *S*'s hypothesizing about *p*, must also be involved in *S*'s acquiring of the belief

⁴⁹ It is also true that the sources of cognitive biases such as that of confirmation bias, availability heuristic, and other judgmental heuristics can result to the biased conclusions without the aid of desire or emotion as in the cases proposed by Martha Knight, "Cognitive and Motivational Bases of Self-Deception: Commentary on Mele's Irrationality," in *Philosophical Psychology*, 1:2 (1998) 179–188. However, they are not cases of SD by the mere fact that they are not motivated.

that p . In other words, if the process deviated or if the belief that p (which is false) is acquired apart from the biasing process, there will be no SD.

An example of deviant causality would be the case of Khloe who favors a belief that a fire that torched a house was caused by electrical short circuit. She sought evidence to support this belief. She consulted Mariah whom she believed is a fan of the theory that burning buildings are usually caused by electrical short-circuit. As it happened, Mariah believed that somebody set the house on fire. She ended up believing Mariah's theory. It was, however, caused by electrical short-circuit. Even if it is false that somebody burned the house, the case is not SD. Khloe's belief is not causally connected to her favored belief. Strictly speaking, SD does not just happen by chance, for we can be misled to believe without being self-deceived. Neither beliefs which are caused by alcohol, hypnosis, torture, accidental blow in the head, can count as cases of SD unless they have a non-deviant connection from the desire and/or emotions. SD might be non-intentional but there remains a connection between self-deceptive belief acquired and its motivational influences.

For the examples of non-deviant causality involved in SD, it would suffice to consider the cases of SD already mentioned in this work. The emperor's belief that he is naked, for instance, is non-deviantly connected to the desire to have worn the magnificent dress as its cause. He is fond of fine suits. This leads him to treat, in a biased way, that a magical suit which only wise men can see is at his disposal. He believes he is wearing one. Adonis' fear of ghost is the non-deviant cause of his belief that there is ghost. Vasyla's jealousy triggers the hypothesis that his wife is unfaithful. It leads him to treat the data in a biased way leading to the acquisition of the false belief that his wife is unfaithful.

This condition about the non-deviant causality assures that to be truly attributable to the subject, it must originate from her as the cause, which is either the self-deceiver's desire or emotions. Cases where there is a deviant causality cannot be properly attributed to her. If twisted cases, as Mele argues, are products of that adaptive or basic desire to avoid the cost of being mistaken, then those cases are violating this condition [4].⁵⁰ For then neither the desire nor emotions towards p is really the immediate cause of the self-deceivers' acquisition of the belief.

Again, in condition [4], I appeal to the role of the motivating influence of desire and emotions in the generation and sustenance of hypothesis in its evaluation as implied in condition [2]. In this view, our LHT theory supports the causal non-deviance of the processes leading to SD. If all goes in accordance with the theory, the hypothesis tester would be seeking for the

⁵⁰ Mele, *ibid.*

confirmation of the proposed hypothesis sustained by desire and/ or emotions. There is thus a continuing causal link from these triggering influences of the hypothesis that p , through its confirmatory biased testing of p , to the belief that p .

{7} The Tension Conditions

The thrust of the conditions [5] and [6] is to highlight the tensions that self-deceiver experiences when he believes against the weight of the evidence available to her. Mele,⁵¹ in formulating his condition (5), is responding to objections against his early set of conditions regarding the lack of tension condition for SD. These theorists have valid reasons to point out the lack of tension in Mele's deflationary account.⁵² Indeed, Mele dismisses that tensions are necessary for SD.⁵³ According to him, there are some cases where one can enter SD in acquiring a belief that p where apparently the self-deceiver does not experience a tension when acquiring the belief that p and yet she can still be regarded as self-deceived. Mele states that:

In some instances of motivationally biased evidence-gathering, for example, people may bring it about that they believe a falsehood, p , when not- 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 $-p$. Such people are naturally deemed self- deceived, other things being equal."⁵⁴

His example of this case is Don, who believed that the rejection of his article was not justified as the referees may have misunderstood his point. When he read the comments after a few days with an impartial mind, he found out that the rejection was warranted. According to Mele:

... he is free of psychic conflict during the process of acquiring the belief that his article was unjustly rejected and while that belief is in place, he is self-deceived and

⁵¹ Mele, "When are we self-deceived?"

⁵² See Robert Audi, "Self-deception vs. self-caused deception: A comment on professor Mele," *Behavioral and Brain Sciences*, 20:1 (1997) 104-104 and W. M. Martin, "Self-Deceiving Intentions," *Behavioral and Brain Sciences*, 20:1 (1997), 122-123.

⁵³ See Mele, "Real Self-Deception" and Mele, *Self-Deception Unmasked*.

⁵⁴ Mele, *Self-deception Unmasked*, 52.

he enters self-deception in acquiring that belief. The same is true of bigots who, without psychic conflict, satisfy my four conditions in acquiring a bigoted belief that p .⁵⁵

He has no qualms, however, with stressing that tensions are important. Thus, in the latest version of his jointly sufficient condition for entering SD in acquiring a belief that p , aside from the tension condition (4): “The body of data possessed by S at the time provides greater warrant for not- p than for p ,⁵⁶ he includes another condition that validates the importance of tensions in SD in his latest article on SD: “ S consciously believes, at the time, that there is a significant chance that not- p .”⁵⁷ Due to those objections above, he conceded that it would be useful for SD to count this condition in his list. But still, he does not make tension a necessary condition. It does, however, help in recognizing supposed cases of SD for tensions usually accompany the phenomenon.

In all this, I agree with Mele when he counted that tension conditions are among the jointly sufficient conditions for entering SD in acquiring a belief that p . However, I have an objection as to the cases of Don and the bigots who acquire beliefs without psychic tension. It seems that there is a psychic tension in Don’s acquisition of the belief that p which is shown by the fact that the news about the rejection of his article has triggered the acquisition of the belief that the rejection was not justified. There could be a psychic tension in Don’s case after all, which can be inferred from the fact that the evidence against the belief that the rejection was not justified was the rejection itself. The tension is between the belief that his paper was unjustly rejected and the evidence that supports that it was justly rejected given by the fact that he received a rejection notice.⁵⁸

Again, LHT theory is useful in elaborating Mele’s condition (5). The theory also pits the evidence for p and not- p against one another. The self-deceiver is somehow aware of the evidence to the contrary, but being motivationally loaded, she is more focused on the ones favoring her hypothesis. Following the design of LHT, we can see the self-deceiver gathering data in favor of her hypothesis until she reaches the acceptance

⁵⁵ *Ibid*, 53.

⁵⁶ *Ibid*.

⁵⁷ Mele “When are we self-deceived?” 11

⁵⁸ It might be argued that some cases of SD where there would be no apparent tensions on the part of self-deceiver is imaginable. Consider the case of chronic or habitual self-deceivers. It might be possible for people to enter SD in acquiring a belief that p without undergoing any psychological tension at all if they do self-deceiving as a habit. Once SD becomes second nature to a person, it also becomes tension-free. However, if it is a habit it might not be a case of SD anymore, but of a serious pathology.

threshold. She may gather the data in a biased, confirmatory way. In doing so, she would need to be open only to the evidence favoring her hypothesis, which would not discount her awareness of the contrary evidence. It is not a tension as strong as the kind of SD that puts belief that p vis-a-vis belief that not- p ; but it is a tension which is also a source of instability for if the motivational force waned in sustaining the confirmatory testing of the hypothesis that p , the hypothesis tester may end up being unbiased.

{8} Lay-Hypothesis Testing Theory and the Psychological Tension Condition

Mele's formulation of his condition (6) is triggered by Scott-Kakures' objection that the former's formulation of the conditions for entering SD are not satisfactory.⁵⁹ Scott-Kakures points out that Mele, in *Self-Deception Unmasked*, neglects an important element in self-deceiver's attitude. In particular, it appears that in his motivated biased belief acquisition account (also of other non-intentionalists' accounts), the self-deceiver is just a passive participant of the processes that surely end to the confirmation of the belief that p . Were it the case, there would be no way of distinguishing SD from wishful thinking. Scott-Kakures illustrates this danger in the case of Bonnie, the feline:

To see the issues in stark form, consider Bonnie, the cat. Like most felines, Bonnie can make fine aural discriminations. She can, for example, distinguish the sounds of the removal of her own medication from the cupboard from the sounds of the removal of other objects—she promptly disappears only when her own medication is removed. Bonnie is also exceedingly fond of her food. She is apt to scamper into the kitchen when she hears her food being opened. She rarely so scampers into the kitchen when some non-cat-food item is being opened. Yet, on occasions upon which Bonnie is very hungry she does certainly appear to mistake non-cat-food sounds for cat food sound

Bonnie comes to have a false belief as a result of a desire Should we say that Bonnie is self-deceived? A deflationist account has scant basis for denying such a characterization.⁶⁰

⁵⁹ Dion Scott-Kakures, "At 'Permanent Risk': Reasoning and Self-Knowledge in Self-Deception, *Philosophy and Phenomenological Research*, 65:3 (2002), 576-603.

⁶⁰ *Ibid*, 578-579.

This is quite a humorous way of considering the limitation of deflationary accounts like those of Mele's earlier formulations. It is because of this objection that the latest formulation of Mele revised his earlier list of four-point jointly sufficient conditions. As a response, he provided condition (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."⁶¹ This condition which is the 6th in his list is set to limit what counts as motivationally biased treatment of data suitable for SD, since some simple motivationally biased misperception cannot count as a case of SD. Accordingly, Mele modifies the previous list to avoid misinterpretation:

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. 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 SD 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 2.⁶²

This justifies his inclusion of the sixth condition that acquiring a belief that p should be a product of reflective, critical reasoning, and S is wrong in regarding that reasoning as properly directed.

The evaluation part of LHT involves steps that entail reflective critical reasoning. The automatic pragmatic considerations of the costs of acceptance or rejection of the data involved such kind of reasoning. The hypothesis tester (being motivated) would always tend to prefer that which is less costly. In this sense, the costlier the belief for the self-deceiver, the more that a reflective critical reasoning is involved, particularly, because so much is at stake if he is mistaken. Scott-Kakures claims that SD also involves failure of self-knowledge, in the sense that the self-deceiver justifies her self-deceptive belief by way of acknowledging that she does not know how she comes to hold such a belief; or that she does not understand at all, but she believes p .⁶³

⁶¹ Mele, "When are we self-deceived?" 12

⁶² *Ibid.*

⁶³ Dion Scott-Kakures, "Can You Succeed in Intentionally Deceiving Yourself?" *Humana Mente Journal of Philosophical Studies*, 5:20 (2012), 17-40.

Whereas condition [6], which states that the self-deceiver must be aware that there is a big chance he is wrong, pertains to the cognitive tension; the psychological tension can be seen more in the condition [7]. Seen within the critical reasoning involved in LHT theory, the psychological as well as the cognitive tensions that the self-deceiver is undergoing are more apparent. While the self-deceiver/hypothesis tester is considering the hypothesis that p , she is experiencing the said tensions.

In the condition [7], we can see that the self-deceiver as a hypothesis tester is not just a passive participant of her own deception. She weighs evidences for and against p . However, because she is more inclined to confirm, rather than reject, the hypothesis that p , she automatically activates the heuristics: for example, she is selective in gathering the data in favor of the preferred hypothesis; she misinterprets data to favor the favored hypothesis; and she attends more to the data supporting the hypothesis. There are evidences that would crop up pointing to the other direction. She may become more aware that she might be wrong which would make the hypothesis testing more intense in confirming that p .

This use of the LHT theory for SD favors the supposition of psychological tension which the self-deceiver might have been undergoing as to the desirability of p in straight cases and undesirability of p in twisted cases. In the straight types the driving force for the self-deceiver is the desire for p to be true against the odds that it might not be the case. It is full of tension and unstable because the desire for p may not be enough to sustain the belief that p . For the twisted cases, it is unlikely that desire for p is the driving force for the belief that p . Rather, it is emotions, such as fear or jealousy, that are pushing for the undesirable belief that p . In these cases, the tension may be more overwhelming. Aside from the question on how such emotions could be pit against the desire, emotions must also be strong enough to sustain the belief that p .⁶⁴

Conclusion

I argued that Mele's list of conditions for entering SD in acquiring a belief that p is more appropriate for cases of straight SD. To accommodate twisted cases of SD, I added a condition to his list. The condition pertains to the motivated generation of hypothesis. This condition not only sets the direction of the actual testing for confirmation of the hypothesis (as in LHT theory) but also guarantees that the processes are motivated. I also explored

⁶⁴ These workings of desire and emotions in LHT theory correspond with the tension brought about by dynamic paradox in intentional SD model. In this model, the tension is undeniably strong as the self-deceiver is expected to have find a way to lie to herself that p even if she believes that not- p .

the significance of the added condition to the rest of the jointly sufficient conditions for entering SD in acquiring a belief that p . Without this added condition, twisted cases could hardly qualify as SD. With the current list of jointly sufficient conditions, all cases of SD are duly accounted for as motivated biased false believing.

Department of Philosophy, Ateneo de Manila University, Philippines

References

- Audi, Robert, "Self-deception, rationalization, and reasons for acting," *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988).
- _____, "Self-deception vs. self-caused deception: A comment on professor Mele," *Behavioral and Brain Sciences*, 20:1 (1997).
- Augustine, "Lying," trans. by Sister Mary Sarah Muldowney, in *Treatises on Various Subjects*, ed. by Roy Deferrari, Vol. 16 of *The Fathers of the Church* (Washington, D.C.: Catholic University of America Press, 1952).
- Bach, Kent, "An Analysis of Self-Deception," in *Philosophy and Phenomenological Research*, 41 (March 1981).
- Bermúdez, José Luis, "Self-Deception, Intentions, and Contradictory Beliefs," in *Analysis*, 60:4 (2000).
- Bok, Sisela, *Lying: Moral Choice in Public and Private Life* (New York: Random House, 1978).
- Davidson, Donald, "Deception and Division," in *The Multiple Self*, ed. by Jon Elster (Cambridge: Cambridge University Press, 1985).
- Demos, Raphael, "Lying to Oneself," in *Journal of Philosophy*, 57:18 (1960).
- Dolcini, Nevia, "The Pragmatics of Self-Deception," in *New Directions in Logic and the Philosophy of Science*, ed. by Laura Feline, Antonio Ledda, and Francesco Paoli (United Kingdom: College Publications, 2016).
- Echano, Mario R., "The Motivating Influences on Self-deception," in *Kritike: Online Journal of Philosophy*, 11:2 (December 2017).
- Fallis, D., "Are Bald-Faced Lies Deceptive After All?" *Ratio*, 28 (2015).
- Friedrich, James, "Primary error detection and minimization (PEDMIN) strategies in social cognition: A reinterpretation of confirmation bias phenomena," in *Psychological Review*, 100:2 (1993).
- Funkhauser, E., "Do the Self-Deceived Get What They Want?" *Pacific Philosophical Quarterly*, 86:3 (2005).
- Gregg, A.P., N. Mahadevan, and C. Sedikides, "The SPOT Effect: People Spontaneously Prefer Their Own Theories," in *The Quarterly Journal of Experimental Psychology*, 70-6 (2017).

- Haselton, M. and D. Buss, "Error Management Theory: New Perspective on Biases in Cross-Sex Mind Reading," *Journal of Personality and Social Psychology*, 78:1 (2000).
- _____, "Error Management Theory and the Evolution of Misbeliefs," *Behavioral and Brain Sciences*, 32:6 (2009).
- Johnston, Mark, "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).
- Klayman, J. and Y. Ha, "Confirmation, Disconfirmation, and Information in Hypothesis Testing," *Psychological Review*, 94, (1987).
- Knight, Martha, "Cognitive and Motivational Bases of Self-Deception: Commentary on Mele's Irrationality," in *Philosophical Psychology*, 1:2 (1998).
- Kunda, Ziva, *Social Cognition: Making Sense of People* (Cambridge Mass.: MIT Press, 1999).
- Lauria, Federico and Delphine Preissmann, "What Does Emotion Teach Us About Self-Deception: Affective Neuroscience in Support of Non-Intentionalism," in *Les Ateliers de L'Ethique*, 13-2 (2018).
- Lazar, Ariela, "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs Under the Influence," in *Mind*, 108: 430 (1999).
- Mahon, J. E., "The Definition of Lying and Deception," in *The Stanford Encyclopedia of Philosophy* (Winter 2016 edition), ed. by Edward N. Zalta, <<https://plato.stanford.edu/archives/win2016/entries/lying-definition/>>.
- Martin, W. M., "Self-Deceiving Intentions," *Behavioral and Brain Sciences*, 20:1 (1997).
- Mele, Alfred, "Approaching Self-Deception: How Robert Audi and I part company" in *Consciousness and Cognition*, 19:3 (2010).
- _____, "Real Self-Deception," in *Behavioral and Brain Sciences*, 20:1 (1997).
- _____, "Self-deception," in *Philosophical Quarterly*, 33 (October 1983).
- _____, "Self-Deception and Delusion" *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation*, ed. by T. Bayne & J. Fernandez, (New York: Psychology Press, 2009).
- _____, *Self-deception Unmasked* (United States: Princeton University Press, 2001).
- _____, "When are we self-deceived?" *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, Francis, "The Goals and Strategies of Self-Deception," in *The Multiple Self*, ed. by Jon Elster (Cambridge: Cambridge University Press, 1985).

- Plato, *Cratylus*, trans. by Benjamin Jowett, in *The Collected Dialogues of Plato*, ed. by Edith Hamilton and Huntington Cairns (Princeton, New Jersey: Princeton University Press, 1961).
- Rorty, Amelie, "The Deceptive Self: Liars, Layers, and Lairs," in *Perspectives on Self-deception*, ed. by Brian McLaughlin and Amelie Rorty (Berkeley: University of California Press, 1988).
- Scott-Kakures, Dion, Scott-Kakures, "At 'Permanent Risk': Reasoning and Self-Knowledge in Self-Deception, *Philosophy and Phenomenological Research*, 65:3 (2002).
- _____, "Can You Succeed in Intentionally Deceiving Yourself?" *Humana Mentis Journal of Philosophical Studies*, 5:20 (2012).
- _____, "Motivated believing: Wishful and Unwelcome," in *Nous*, 34:3 (2000).
- Talbott, William, "Intentional Self-Deception in a Single Coherent Self," in *Philosophy and Phenomenological Research*, 55:1 (1995).
- Tolstoy, Leo, *The Kreutzer Sonata* (Greenwich, CN: Fawcett Publications, 1960).
- Trope, Y. and A. Liberman, "Social Hypothesis Testing: Cognitive and Motivational Mechanisms," in *Social Psychology: Handbook of Basic Principles*, ed. by E. Higgins and A. Kruglanski (New York: Guilford Press, 1996).