ANNA ELISABETTA GALEOTTI
Università del Piemonte Orientale, Vercelli
elisabetta.galeotti@uniupo.it

STRAIGHT AND TWISTED SELF-DECEPTION

abstract

The paper analyzes the two types of self-deception, usually labeled straight and twisted self-deception. In straight cases the self-deceptive belief coincides with the subject's desire. In twisted cases, by contrast, the self-deceptive belief opposes the subject's desire as in the example of Othello's conviction of Desdemona's infidelity. Are both these contrasting types of deceptive beliefs cases of SD? The argument of this paper shall answer this question in the positive, yet in different way from the unitary explanation of straight and twisted SD proposed by Alfred Mele. The causal account of SD claims to provide a unitary and simple explanation for both straight and twisted SD, and considers such a unitary explanation as a specific virtue of the causal view. Within the same causal model, the difference between straight and twisted self-deception is explained by a difference in the motivational state that in twisted cases is dominated by emotions. The paper will critically examine this claim, and advance an alternative explanation based on a different view of self-deception where emotions play a role alongside wishes both in straight and in twisted case.

1. Introductory remarks

In this paper, I intend to face the issue of two types of self-deception (SD) usually labeled straight and twisted self-deception. I take that SD is believing that P against the available evidence and on the basis of one's wish that P. In straight cases, the self-deceptive belief coincides with the subject's desire. For example Swann, the hero of the first book of Proust's *La Recherche,* desperately wishes that his lover Odette is faithful to him; hence he discounts all the evidence suggesting Odette's infidelity so as to go on believing her faithful[1]. In twisted cases, by contrast, the self-deceptive belief opposes the subject's desire as in the example of Othello who, despite wanting Desdemona faithful to him, becomes convinced of Desdemona's infidelity in the absence of the appropriate evidence. Can we consider both these contrasting types of deceptive beliefs as cases of SD?

In the philosophical discussion, SD is accounted for by two opposite models: the intentional, according to which SD is a doing of the subject, though non-transparently, and the causal, according to which SD is a mental event caused by a motivational state triggering cognitive biases. Twisted SD represents one of the several weaknesses of the intentional model, as noted by the eminent representative of the causal account, Alfred Mele (2001). Intentionalists view SD as a non-transparent plan of the subject, vicariously serving her goal to believe what she wishes to be the case (Davidson 1985, Talbott 1995, Bermudez 2000). In twisted cases, however, the subject ends up believing what she does not wish, hence the purposive nature of her deceptive belief is undermined. Why should anyone engage in such a self-sabotaging plan, though unconsciously? Within the intentional account, twisted SD is a problem, requiring a different kind of account.

By contrast, the deflationary causal-motivationalist account of SD claims to provide a unitary and simple explanation for both straight and twisted SD, and considers such a unitary explanation as a specific virtue of the causal view (Lazar 1997, 1998; Mele 2001, Scott-Kakures 2002).

I will critically examine this claim, and argue that the unitary model provided by the causal account pays the price of its deflationary approach and is much less unitary than is claimed. I shall propose an alternative explanation, which makes sense of the difference between straight and twisted cases; my alternative explanation is based on what I have labeled the

---

1 Both Swann and Othello have more complicated attitudes, feelings and beliefs, but here I consider them only in sketchy manner as types of straight and twisted SD

'Invisible Hand model of SD' (Galeotti 2012) and is carried out by means of a more detailed analysis of the composition of the motivational state.

My argument proceeds as follows: I shall start by presenting the deflationary account as provided by Alfred Mele (2001), and pointing out its weaknesses. Then, I shall briefly outline my alternative account of SD, patterned after the invisible hand model, and turn to the unitary explanation of straight and twisted cases. I shall focus my critical analysis on unpacking the black box of the "motivational state." In this way, I shall be able to understand the similar mechanism in both straight and twisted cases as well as to explain their difference. The difference between the two SD types lies in the different emotional response to the threat of P not being true, each leading to a specific cognitive option and to related biases.

Mele lists the following as jointly sufficient conditions for SD:

2. The deflationary account critically examined

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 *non-p* than for *p*.

I hold that these four conditions are actually insufficient a) to characterize SD as a specific phenomenon in the range of motivated irrationality; b) to explain the selective working of SD, which is not the general cognitive shortcut taken all the time we meet with unwelcome evidence; c) to provide a satisfactory unitary model for both straight and twisted cases. My argument will concentrate on the third issue, but will advance some remarks also on the others.

The four sufficient conditions state that a belief is self-deceptive if 1) it is false; 2) it is causally produced by cognitive biases; 3) triggered by a (yet undefined) motivational state; 4) when the subject possesses evidence contrary to the belief in question. Condition (4) does not specify if the data "possessed by S" are merely available to her, in front of her eyes, or are actually appraised by S. Moreover, the order of the four conditions is such that the intrinsic link with condition (2), concerning the biased treatment of data, is obscured. Actually SD is initiated because S has appraised data contrary to, hence undermining, the truth of S's wish that P. S's emotional response to this threat leads her to linger on her thoughts which become easily affected by biases, ending up in the belief that P, which is false but corresponds to S's anxious wish.

The two points of a) the appraisal of contrary evidence and b) the consequent priming of the biased treatment of data, and in that order, are to my mind crucial to account for SD as a specific instance of motivated irrationality. If the subject believes what she wishes, and if it contingently turns out that what she wishes happens to run against available data, then SD is just a case of wishful thinking gone badly. If instead SD is the motivationally distorted response to the threat of the negative evidence that the subject perceives or suspects as running against her wish that P, then it is a distinct phenomenon in the range of motivated irrationality. In the first case, SD is not sufficiently marked off doxastic states such as wishful-thinking and illusions. The distinctiveness of SD is simply entrusted to condition (1), that P is false, which however is just definitional, and does not affect (2), that is the production of the self-deceptive belief. In order to preserve the specificity of SD I hold *firstly* that condition (4) must be revised as "S appraises of data contrary to her wish that P, making her suspect that P is under threat"; and *secondly* the ordering position of the four conditions must be changed to make clear that the biased treatment of data is primed by the threatening suspicion aroused by the appraisal of negative data. Moreover, the distinctiveness of SD further requires that the condition "S has a wish that P" be added to the list. A distinctive feature of self-deceptive

beliefs compared to other motivated false beliefs, such as stubborn ones, is that the content of self-deceptive belief is thematically linked to the motivating wish (Lynch 2013).

If the SD process starts as the reconsideration of the negative evidence concerning the wish that P by a subject under a motivational state, then the process cannot simply be described as the causal triggering of motivation on biases which in turn causally affect the subject's reasoning. I agree that the product of SD is unintended, but I hold that the mental process producing SD is intentionally entered into as a response to the suspicion of ~P. Thinking matters over, when a threat to one's own crucial wish that P, is not epistemically faulty; but the emotionally loaded wish that P is likely to unleash cognitive biases in considering the available evidence, bringing about the self-deceptive belief. Briefly, I propose an explanation of SD patterned after the 'invisible hand model' to explain SD. The model, advanced for the explanation of seemingly purposeful social phenomena (Nozick 1977), provides a neat account of the coexistence of intentional doing, causal composition, and unintentional outcome, for it explains a certain beneficial effect as "the result of human action, but not of human design" (Hayek 1952). Similarly, in SD the intentionality of process must be sharply distinguished from the non-intentionality of outcome, despite the apparently self-serving result. The invisible hand model can also make sense of the apparent purposefulness of the SD product, which, though it is not a goal of the self-deceiver, however is not simply contingent. Arriving at the false belief that P is the effect of the subject's biasing, but stopping at P is her intentional doing. As I shall show late, also twisted cases, in a way, serve a goal of the subject though definitely not brought about with that intention.

3. Desires in Straight and Twisted Cases

Turning now to straight and twisted case, the analysis of motivation, as suggested by the causal model, raises the following controversial issues:

1) Does the content of the motivational state change in the case of twisted SD compared to straight SD? If it does, is the explanation of both straight and twisted cases still unified? (Mele 2001, Lazar 1999, Scott-Kakures 2002).

2) Is the operative desire for producing either straight or twisted SD the desire that P be the case or the desire to believe that P? (Nelkin 2002)

3) Should there be a match between the operative desire and the content of self-deceptive belief or is the operative desire content-unrestricted compared to the self-deceptive belief? (Nelkin 2002, 2012).

Concerning point (3), I agree with Nelkin that there should be a match between the operative desire and the content of the self-deceptive belief, otherwise a) it is not clear that it is a case of SD, for the self-deceptive belief must be thematically linked with the original motivation; b) methodologically, the shift looks an ad hoc adjustment. Think of the example of Swann: his self-deceptive belief that Odette is faithful to him corresponds to his desire. Such correspondence allows any external observer to suspect SD, given that the resulting belief that P fulfills Swann's desire and, at the same time, runs against the evidence available to him. Next, consider the example of Othello. Following Mele, Othello's original desire of Desdemona's fidelity is not the operative one producing the (false) adverse belief that she is guilty. According to Mele, here the original desire is trumped by another—the desire of not being fooled—which becomes operative in triggering data biasing and producing the false belief of Desdemona's infidelity. Thus, the operative desire has no content-match with the resulting self-deceptive belief, and in Nelkin's terms it is content-unrestricted.

Point (2) is raised by Dana Nelkin in her attempt to find an alternative solution to the content-restrictedness issue for twisted SD. She thinks that it is possible to keep unitary the causal explanation in both straight and twisted case if what the subject wishes is not P, but rather *to believe that P*. Swann wishes *to believe Odette faithful*, in order to keep his peace of mind, while,

symmetrically, Othello wishes *to believe Desdemona unfaithful*, in order to avoid being fooled. Under this description, in both cases there is a match between the operative desire and the self-deceptive belief. This apparently neat solution however leaves much unanswered. Does the self-deceiver wish to believe that P, no matter what, or rather wish that the belief that P be true and consequently be believed? Why should anyone want to believe that P no matter what? The self-deceiver is epistemically confused, but is not incoherent or crazy. Those who think that the operative desire is *to believe that P* assume that there has been an unconscious trade-off between the wish that P be the case and the wish to believe P no matter what. True, at the end of the self-deceptive process, the subject precisely gets the belief that P devoid of its truth-value. But what the subject eventually gets does not make such belief the primary object of her desire, nor of her unconscious aim. Actually the supposed trade-off between states-of-the world and beliefs would make the pursuit of the false belief that P a crazy intent indeed (Lazar 1996). Nelkin's proposed solution for keeping unitary the explanation of straight as well as twisted cases must be rejected.

Coming now to point (1) concerning the composition of the motivational state, Mele holds that there is a difference between either straight or twisted SD (2002, 111ff). In the former, the driving component is desire, in the latter is emotions; other scholars supporting the causal account instead point to different types of emotions in either SD types (Lazar 1999). In general, in the causal account, the specific role of emotions and desires is not clearly set apart. Mele, on the one hand, attributes the difference between straight and twisted cases to the different causal role of desires and emotions respectively; on the other hand, illustrating the case of the jealous husband, he also singles out an operative desire triggering the biasing as well, the one criticized by Nelkin for not being content-restricted. Ultimately, it is not clear whether the twisted cases are caused by emotions or by the operative desire or by both, being the emotions the cause of the desire's switch. The lack of a clear specification of the motivational state weakens the causal account of the twisted cases, and, besides, attributing the causal factor in one case to desires and in the other to emotions makes the account less unitary than claimed.

I argue instead that the components of the motivational state are the same in both straight and twisted SD, and, more precisely, that the motivational state giving rise to SD processes of all sorts comprises the desire that P and the emotions of fear and anxiety loading the desire that P and making it "rebellious" (Pears 1994). How to account of such composition of the motivational state? Given the desire that P—often already present, albeit dormant in the subject—two elements contribute to engender the motivational state favorable to start a self-deceptive belief formation process. *The first* is the appraisal of the evidence threatening the truth of P. Before that moment, Swann or Othello did not think of their lovers' fidelity or infidelity. It is only when Odette fails to meet Swann at the Verdurin's salon one night, or when Iago insinuates doubts about Desdemona, that jealousy is aroused. Since the emotional arousal proceeds from a pre-attentive appraisal and not from a proper epistemic attendance to and processing of, the negative evidence, it may well be the case that such appraisal is actually groundless. This is important for the account of twisted SD, such as Othello-like jealousy. In any case, the suspicion about one's lover's fidelity arouses emotions of fear and anxiety in the subject that make the wish of her fidelity the anxious focus of the agent's desire-set. In other words, the desire that P is "turned on" emotionally by the threat and the suspicion about P not being true. At this juncture *the second* element becomes paramount: the subject is placed in circumstances such that acting in order to produce the state of affairs where P is made true is perceived as beyond her or his control. This is a typical circumstance for SD to start. When the subject is powerless about bringing about the desired state of affairs by a plan of action, he lingers in his thought. The fact that the fulfilment of his desire is beyond his reach, has the

effect of sinking the costs of inaccuracy in acquiring and processing data relevant to the truth of P. A properly diagnostic attitude will not mend the fact of Odette's infidelity for Swann.

4. Costs of Inaccuracy and Selective Vigilance

This point is nicely illustrated by Robert Jervis (Jervis 1976), who argues that the level of accuracy in gathering and processing information correlates with real-life incentives and prices for one's welfare. The costs of inaccuracy are irrelevant for Swann, for no attitude, vigilant or otherwise, will change Odette's conduct towards him. In similar fashion, Mele resorts to the notion idea of a focal error to be avoided (Friedrich 1993, Trobe and Liberman 1996). In case, no matter what the subject believes or does, he cannot bring about what he most wishes, the focal error to be avoided is not a dangerous outcome, which is rational to expect in case of incorrect information, but rather is exposing oneself to a painful truth *unnecessarily*, for changing the state of the world is beyond one's reach. Compared to Mele's explanation, Jervis adds a further non-cognitive component to the understanding of selective vigilance and the lack of it, namely the level of anxiety and stress concerning evidence processing. Low and high level of anxiety typically induce less accuracy and vigilance than a medium level of stress. But while low anxiety leads the subject to rely on routine and traditional patterns of conduct, high anxiety and stress tend to engender "defensive avoidance," that is a blocking of the negative information and reliance on a false soothing belief, which is precisely a form of SD. Clearly, this explanation holds only under the condition that the subject is an agent, though confused, and not a victim of the emotional state that triggers the biasing behind his back. Only an agent can sense, perceive, or suspect the need for vigilance.

To sum up: Odette's love and fidelity is beyond Swann's control. Hence the cost for inaccuracy sinks, since Swann has no expectation of changing Odette's mind through more accurate information. Besides, a positive correlation between high anxiety and defensive avoidance has also been shown: therefore, Swann not only cannot gain anything from a more accurate information of Odette's bearings, but his emotional state also tends to cause a cognitive attitude of "defensive avoidance", for, under his unfortunate circumstances, the focal error to be avoided is disbelieving Odette faithfully unnecessarily. Consequently, the threshold of evidence Swann requires to revise his belief about Odette's fidelity is disproportionally high. It is not a coincidence that typical examples of SD have to do with questions which are at the same time *crucial* for the subject's *well-being* and *beyond his control*: marital infidelity, fatal illness, and children's problems of addiction or criminal behavior. The first circumstance of crucial desires heightens anxiety, while the second makes the costs of accuracy sink.

These same two circumstances occur in twisted SD as well: a state of affairs crucial for the subject's well-being is threatened by the suspicion of contrary evidence that is beyond the subject's control. But the same circumstances may engender different forms of bias depending on different emotional and cognitive responses.

5. Revising the motivational state

Coming back now to the motivational state, I shall propose the following unpacking:
1) The operative desire is the desire that P and not the desire to believe that P.
2) The operative desire is a top priority in S's preference ranking.
3) The costs of inaccuracy in acquiring and processing data relevant to the knowledge that P is low or irrelevant for S.
4) The desire that P is emotionally loaded by fear and anxiety.
5) These two emotions are aroused by the appraisal or misappraisal of negative evidence threatening the truth of P.

The three issues left unsolved in the causal account, namely specificity, selectivity, and the unitary explanation of both straight and twisted SD, can now find a satisfactory response.

Briefly, (5) and (4) within the invisible hand model explain the specificity of SD. The desire that P is emotionally loaded by fear and anxiety produced by the appraisal of negative evidence as to the truth that P. S, anxiously suspecting that P is under threat, and powerless to counter the threat, starts thinking over the matter, and becomes prey to cognitive biases under the influence of her emotionally loaded wish that P. SD is entered only under condition (3), concerning the sinking of accuracy costs in the circumstances, and condition (4) inducing defensive avoidance. Thus, (3) and (4) jointly explain the selectivity of SD, but only under the condition that SD is intentionally entered. If by contrast, the subject is the victim of the motivational state getting a causal grip on cognition, activating biases, and producing the motivationally false P, then he can hardly be sensitive to costs of accuracy/inaccuracy and threshold of vigilance. Moreover, the biases' activation is secured by the motivational state ex ante, hence the costs of accuracy/inaccuracy seem redundant for starting SD.

Let's turn now to the unitary explanation of twisted and straight SD. The supporters of the deflationary account say that in either SD type, a motivational state triggers cognitive biases ending up in the fictional belief that P, either favorable or adverse. (Scott-Kakures 2000: 354). In twisted cases it is emotions rather than desires that switch on the biasing, and that explains the different result. Mele adds to this explanation a switch in the operative desire that is no more the desire that P (wife faithful) but the desire that Q (not being fooled). Thus this unitary model implies two unjustified shifts in twisted cases compared to straight ones: a) the emotional component takes the lead and causes the bias ending up in the false adverse belief; b) there is a shift in the operative desire as well, from the desire that P to the desire that Q. Both shifts are unexplained, and taken together they explain too much. If there is a shift in the operative desire from P to Q, then there is no need to suppose that the causal job is done by emotions, for the desire that Q is acting precisely like the desire that P in straight cases. Yet the resulting adverse belief ~P is not matching the operative desire that Q, thus giving rise to the issue of content-unrestrictiveness. If instead, the causal job is done by emotions, there is no need to shift the operative desire. It is still unexplained why S (wrongly) suspecting that P is under threat, becomes prey to the biases unleashed by fear and produces the false adverse belief ~P, and S[1] (rightly) suspecting that P is under threat, let the desire that P to switch on biases producing the false favorable belief that P. The difference cannot lie in the fact that suspicion is right in straight cases and wrong in twisted ones, for it being right or wrong is an external condition and is apparently independent from the subject's motivational and epistemic setting.

6. Alternative explanation for both straight and twisted SD

As argued above, the motivational state has the same content for either case of SD. That S is moved by ungrounded emotions does not change how the SD process is started. In both cases, S is not (or believes not to be) in a position to act in order to undo the threat and bring about the state of affairs ex-ante. Like in most SD cases, S faces matters crucial to her well-being, yet beyond her control, actually lowering the costs of inaccuracy. The response is a thinking process affected by biases ending up in the deceptive belief. This is the common model operating in both cases; let us now consider the difference between straight and twisted SD. In straight SD, self-deceivers usually display confirmatory bias that is the mental habit to look for confirmation of one's hypothesis, while epistemic rationality would recommend falsification as the sound procedure. Confirmatory bias leads self-deceivers to ignore or discount the negative evidence and to focus instead on the positive data, ending up with the false but favorable belief that P. In twisted cases, instead, S imagines what she most fears in order to face the worst possible case. In principle, such a strategy is not faulty: it is justified by the insurance logic that is a strongly risk-averse approach of the kind: "let's assume the worst, and work out a maximin response in terms of attitudes, beliefs and actions to such a

scenario". The problem with this strategy, however, is that worst case scenarios become easily affected by a special bias—probability neglect that makes the reasoning faulty. In principle, worst-case scenarios are improbable events, representing only one extreme of a whole range of possibilities, yet it seems that once the scenario has been conjured up, the low probability is discounted, and the subject's fear and anxiety are reassured only by conclusive evidence showing the scenario false. Thus while S lowers the threshold of evidence to go on believing that P, S^1 heightens the threshold of evidence required to disbelieve ~P .The operative desire that P is matched by the resulting self-deceptive belief that ~P, though negatively, so as to be thematic, thus satisfying a commonly acknowledged requirement for SD.

The general explanatory model is the same in both cases: an emotional arousal following the appraisal, or the misappraisal, of negative evidence makes the wish that P the anxious focus of the subject; the active response to bring about P is foreclosed, hence the subject starts to think instead. The reasoning, influenced by the wish, attempts either at confirming the belief that P despite contrary evidence, or at disconfirming ~P, in the absence of the relevant evidence. In the first case, the reasoning is affected by confirmatory bias, and in the latter by probability neglect. In neither case the outcome is intended though produced by intentional thinking. As much as the confirming strategy is the common mode of the lay hypothesis testing and is generally pragmatically reasonable (though epistemologically imperfect) similarly, the counter-factual mode of reasoning implied by the worst-case-scenario is, as said, at the basis of the insurance logic. In both types, normal pragmatic modes of reasoning are distorted by the anxious fear concerning P. As a result, in straight SD the false belief that P is confirmed, and in twisted SD, the opposite, but likewise false, belief ~P is produced. Admittedly, twisted SD does not serve the practical goal of dispelling S's worries, but it may perversely relieve a certain kind of anxiety, still fulfilling a purpose for especially anxious subjects. For such subjects usually display a preference for certainty over uncertainty even of unwelcome truths; consequently, they may be miserable, but will also feel a certain relief from the anxiety of suspecting the worst, and a (deceptive) sense of control.

7. Conclusions

Twisted SD cases pose a challenge for the intentional view of SD, according to which the self-deceptive belief fulfills a goal of the subject. Purposefulness cannot be assumed in the case of twisted SD, given that the self-deceptive belief opposes the subject's desires. It may actually be the case that extremely anxious subjects have a preference for unpleasant certainty over uncertainty. Yet, within the intentional framework it remains unclear why such overanxious subjects would not try to relieve their anxiety by believing something soothing, thus overcoming uncertainty and unwelcoming evidence at the same time. Causal accounts apparently present a clear advantage with reference to the issue of twisted cases. The latter need not be assumed as the outcome of special perversity. Twisted cases are a type of SD, produced in the same way as straight SD: they are the result of a motivational state causally switching on cognitive biases resulting in a false belief, no matter if favorable or adverse.

The appeal of such a unitary and simple account of both types of SD is clear. However, if this account is duly analyzed, then it turns out to be less unitary than claimed, for the causal switch is not the same in either case, and moreover, the operative desire in the twisted case is content-unrestricted.

My alternative invisible hand account provides a general model for either straight or twisted cases, which can be articulated in different reasoning strategies thus making sense of the two opposing outcomes as well. The invisible hand account holds that the SD process is entered intentionally and legitimately too: it is the response of an agent who has met with negative evidence concerning one of her desires, and is powerless to counter the threat. Under the circumstances, the agent is anxious and cannot act rationally, resolving her worries by fixing

the problem. The agent lingers in thinking about the problem, and thinking is something that she does intentionally. Yet the fact of her powerlessness, along with her anxiety, lowers costs of accuracy, manipulating the threshold of evidence required to either disbelieve or believe that P is true. The result is the relaxation of her usual epistemic standards without her noticing. This is the common track of SD of either type, made up of intentional steps and non-intentional twists; what changes is the cognitive strategy. If the subject is testing the truth of P, and in so doing, is caught in by confirmatory bias, we have straight SD, with the deceptive belief matching the agent's desire. If the subject conjures up a worst-case scenario, she falls victim to probability neglect, and without a final disproof she will be caught by the false and adverse belief that ~P. The choice of the strategy is intentional, but the related bias is not, just as the outcome is unintended.

To conclude: the same invisible hand model explains the self-deceptive production of both favorable and adverse beliefs, without the need to shift the causal components from desires to emotions, and without the need to shift the operative desire. The operative desire is always a) emotionally overloaded; b) the same that originates the worries of the subject; c) content-restricted in either case, though in twisted case it is in the negative form. If the product of twisted SD is an adverse belief, feeding worries and fears of the subject, it can still be said that in a special sense the belief serves a goal of the subject by dispelling her uncertainty over what to believe.

REFERENCES

Bermudez, J. L. (2000), "Self-Deception, Intention and Contradictory Beliefs" *Analysis*, 60, pp. 309-319;

Davidson, D. (1985), "Deception and Division", In LePore, E., and McLaughlin, B. (eds.), *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Basil Blackwell, Oxford, pp. 138-48;

Friedrich, J. (1993), "Primary Error Detection and Minimization (PEDMIN) Strategies and Social Cognition. A reinterpretation of Confirmation Bias Phenomenon", *Psychological Review,* 100, pp. 298-319;

Galeotti, A. E. (2012), "Self-Deception: Intentional Plan or Mental Event?" *Humana Mente*, 20, pp. 41-66;

Hayek, F. A. (1952), *The Counter-Revolution of Science*, The Free Press of Glencoe, Collier-Macmillan Limited, London;

Jervis, R. (1976), *Perception and Misperception in International Politics*, Princeton University Press, Princeton;

Lazar, A. (1997), "Self-Deception and the Desire to Believe", *The Behavioral and Brain Sciences*, 20, pp. 119-20;

Lazar, A. (1999), "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs 'Under the Influence'", *Mind*, 108, pp. 265-290;

Lynch K. (2013), "Self-Deception and Stubborn-Beliefs" *Erkenntis*, 78, pp. 1337-45;

Mele, A. (1999), "Twisted Self Deception", *Philosophical Psychology*, 12, pp. 117-137;

Mele, A. (2002), *Self-Deception Unmasked*, Princeton University Press, Princeton;

Neklin, D. (2002), "Self-Deception, Motivation and the Desire to Believe", *Pacific Philosophical Quaterly,* 83, pp. 384-406;

Nelkin, D. (2012), "Responsibility and Self-Deception: A Framework", *Humana Mente*, 20, pp. 117-39;

Nozick, R. (1977), "On Austrian Methodology", *Synthese*, 36, pp. 353-392;

Pears, D. (1984), *Motivated Irrationality*, Oxford University Press, Oxford;

Scott-Kakures, D. (1996), "Self-Deception and Internal Irrationality", *Philosophy and Phenomenological Research*, 41, pp. 31-56;

Scott-Kakures, D. (2000), "Motivated Believing: Wishful and Unwelcome", *Nous*, 34, pp. 348-75;

Scott-Kakures, D. (2002), "At Permanent Risk: Reasoning and Self-Knowledge in Self-Deception", *Philosophy and Phenomenological Research*, 65, pp. 576-609;

Talbott, W. J. (1995), "Intentional SD in a Single, Coherent Self", *Philosophy and Phenomenological Research*, 55, pp. 27-74;

Trobe, Y., and Liberman C. A. (1996), "Social Hypothesis Testing. Cognitive and Motivational Mechanism" In Higgins, E., and Kruglansky, A. (eds.), *Social Psychology: Handbook of Basic Principles*, Guildof Press, New York, pp. 239–270.