Deceiving Oneself Or Self-Deceived? On the Formation of Beliefs “Under the Influence”

ARIELA LAZAR

How does a subject who is competent to detect the irrationality of a belief that \( p \), form her belief against weighty or even conclusive evidence to the contrary? The phenomenon of self-deception threatens a widely shared view of beliefs according to which they do not regularly correspond to emotions and evaluative attitudes. Accordingly, the most popular answer to this question is that the belief formed in self-deception is caused by an intention to form that belief. On this view, the state of self-deception is taken to be a calculated outcome involving a person’s intentional manipulation of her own thoughts. I argue that this answer is false and forms an impediment towards making sense of self-deception. I show that, contrary to philosophical prejudice, emotions and desires exert vast and systematic effects on the formation of beliefs. In this, and other, sections of the article, the results of experimental work are brought forward. Self-deception is portrayed here as resembling numerous instances of belief formation which are regularly affected by motivational factors. I argue that self-deceptive beliefs are direct expressions of the subject’s wishes, fears and hopes. Qua beliefs which mostly correspond to such factors (rather than to evidence), self-deceptive states are a kind of fantasy.

1. Introduction

Joan’s husband changed his demeanor recently. He no longer wishes to spend time with her, is evasive when describing his engagements, tends to arrive late from work and has many unexplained expenses. Joan’s friends alert her to the fact that her husband is being unfaithful, but in spite of their comments and the changes in Bill’s behavior, Joan is convinced that he is not. Joan, it seems, is self-deceived about her husband’s actions.

Cases such as Joan’s are both familiar and puzzling. This essay focuses on what is, no doubt, the central question that arises vis-à-vis such cases: How does a subject who is competent to detect the irrationality of a belief that \( p \), form her belief against weighty or even conclusive evidence to the contrary? In attempting to answer this question, I first examine the most popular answer, that the belief formed in self-deception is caused by an intention to form that belief. I shall argue in §2 that this answer is wrong and forms an obstacle to understanding self-deception. The remainder of this essay lays out the elements which figure in the account I shall
endorse. §3 describes the vast and systematic effects of emotions and desires on the formation of beliefs. These effects are not explainable by appeal to the subjects’ intentions. §4 explores the relation between imagining and believing and lays out the basic concepts within which self-deception is to be understood.

On my view, self-deceptive beliefs are direct expressions of the subject’s wishes, fears and hopes. Qua beliefs which mostly correspond to such factors (rather than to evidence), self-deceptive states may be described as fantasies. Even so, these beliefs affect the subject’s behavior just like normal beliefs, by figuring in practical and theoretical reasoning. In this essay, self-deception is shown to be a consequence of very common processes by which desires and emotions regularly affect the formation of beliefs. These effects are exhibited in a wide variety of instances and are not limited to cases of self-deception; nor are they mediated by practical reasoning. The exploration of self-deception in this article is viewed, in part, as a means towards revealing this feature of beliefs. In this way, I hope to discredit a philosophical prejudice according to which beliefs do not regularly correspond to desires and emotions.

Self-deception is a species of irrational belief formation. Cases of irrational belief formation are of two kinds: those in which a motivational factor is causally effective in bringing about the belief—cases of motivated irrational belief formation—and those in which such a factor is not at play—“cold” cases. Self-deception is a species of motivated irrationality: it is generally taken to consist of a process which results in the subject’s holding an irrational belief the formation of which is partially due to the presence of a desire or some other evaluative attitude. The formation of the irrational belief is not due to the subject’s intellectual incompetence in the following sense: detecting the irrationality of the belief in question does not require intellectual abilities that exceed those of the subject. Similarly, self-deception is not due merely to the presence of a cognitive bias. A simple example of a cognitive bias is the overvaluing of evidence that is encountered early on in the process of information gathering.

1The phrase “irrational belief formation” is ambiguous. It may be interpreted as characterizing a belief as irrational or it may be interpreted as characterizing as irrational the process that leads to the formation of a belief. In using this phrase, I intend that it be understood in the first sense. I say nothing here about whether, and in what way, processes through which beliefs are formed may be assessed as rational or irrational.

2David Pears (1984) stresses this distinction.

3The term “self-deception” is regularly used both in the context of discussing the process by which the irrational belief is formed as well as its end state (the state of being self-deceived). Below, I mostly use “self-deception” to designate the process and use more specific phrases (for example, “the end product of self-deception”, “the self-deceptive state”) to designate the state of being self-deceived.
Cognitive biases are persistent and highly prevalent patterns of biased reasoning. They are exhibited regardless of subject matter. In contrast, self-deception is thematic: the content of the irrational belief is relevant to the explanation of its formation. Recall Joan’s irrational belief that her husband is not having an affair. Joan’s failure of “putting together” properly the pieces of the evidence is due in part to a complex of emotions and desires concerning her husband and her relationship with him (for example, her desire that her husband be faithful to her, her love for him). If cognitive biases are to play a role in explaining self-deception, the explanation must be complemented by showing how the content sensitivity (or the apparent content sensitivity) of self-deception is generated. That self-deception is not due to a mere cognitive bias is also indicated by the way in which self-deceived persons respond to criticism. Self-deceived subjects are more likely to resist the correction than competent subjects in the grip of a cognitive bias.

Cases of motivated irrational belief formation may be grouped according to the extent of irrationality involved (compare a case of a person who holds a belief which accords with the examined evidence but refrains from exploring further seemingly relevant information, with that of a person who has conclusive evidence for \( p \) and holds not-\( p \)). It is debated which cases ought to count as cases of self-deception. The most extreme cases involve the subject’s initial recognition that the evidence at his disposal supports a certain belief (\( p \)) and yet forms its negation (not-\( p \)). The account of self-deception that is endorsed here implies that such an initial recognition on the part of the subject is not a necessary feature of standard cases of self-deception. But I will not attempt to settle the matter of what degree of irrationality is required in order that a case count as self-deception. Everyday use of “self-deception” is not explored in this essay. Rather, this essay focuses on uncovering the process that leads to the formation of the irrational belief. In this spirit, I withdraw from such debates as that concerning whether, on the basis of common use, it should be claimed that the belief formed in self-deception must be false. For the same reason, I will not engage in the debate over whether “self-deception” applies to cases that involve a wish’s directly causing the formation of a belief (widely known as wishful thinking).

The literature in this area is vast. Examples include Tversky and Kahneman (1972, 1973, 1974, and 1981). Nisbett and Ross (1980) present a useful survey of a variety of findings in this area.

Clearly, a person who merely made a mistake may resist correction (for example, to avoid public humiliation). However, a self-deceived agent resists correction even in the absence of unwelcome social consequences.

On whether the end product of self-deception must consist in a false belief, see Mele (1987, Ch. 9) and Pears (1984, Ch. 3). On whether self-deception is a kind of wishful thinking, see Pears (1984, Ch. 2), Davidson (1986) and Johnston (1988).
Finally, it ought to be acknowledged that an investigation of self-deception faces a special challenge: one is not aware of one’s predicament at the time during which one is self-deceived. Typically, if one is aware of any self-deception, it is another person’s, or one’s own self-deception in the past. Lack of awareness of one’s current self-deception rules out a phenomenological investigation of this phenomenon. This feature sets apart self-deception from the comparable phenomenon of acting against one’s own better judgment. There is no particular difficulty, in the latter case, in an agent’s being aware of her acting against a better course of action. Indeed, the clearest cases of akrasia involve the agent’s full awareness of this fact.

2. The failure of the intentionalist account

2.1 The intentionalist account

In the philosophical literature, self-deception is often described as an action. It is said to be something in which the subject of self-deception engages for a reason, or an action that is performed with a special kind of intention.\(^7\) This philosophical stance has the advantage of corresponding to one way in which self-deception is discussed in non-theoretical contexts: we often say of someone whom we suspect to be self-deceived that he is deceiving himself. Saying of someone that he is deceiving himself may be construed naturally as an attribution of an action to the subject of self-deception. The act of deception (in the case of other-deception) involves an intention to cause the formation, in another person, of a belief which is deemed false by the deceiver. Accordingly, the most popular philosophical account portrays self-deception as involving an intention to form a belief which the subject deems false (or discredited) at the time in which he undertakes the project of self-deception.

The main reason for the popularity of this philosophical approach is the following. The self-deceived subject is highly irrational—she forms a belief that is undermined by the weight of the evidence even to the extent of the evidence being overwhelming in support of its negation. At the same time, the presence of the irrational belief that is formed in self-...

\(^7\) Those who hold that self-deception stems from an intention to form a belief include Davidson (1986), Pears (1984), Rorty (1988), Talbot (1955) and Whisner (1992). Psychologists often endorse this view of self-deception (Quattrone and Tversky 1984, and Gur and Sackheim 1979). Mele (1987, 1997) purports to reject this view but offers a relatively similar account which is susceptible to some of the objections below. In addition, Mele’s account faces another crucial objection (for a critique of Mele’s view, see Lazar 1997). Gardner (1993) seems to realize the limitations of this view but does not make a direct attempt to undermine it.
Deceiving Oneself Or Self-Deceived? 269

decception often corresponds to a goal (or goals) of the subject. The irrational belief may alleviate anxiety or boost the subject’s self-image, while the rational belief threatens the fulfillment of these goals. It is suggested then that the irrational belief is acquired in order to attain a goal that is frustrated by the presence of the rational belief. Since, by assumption, the self-deceived subject is competent to detect the irrationality of her belief, it may seem that this is one of very few available explanations of its formation. The view that the irrational belief is formed intentionally, as a means of attaining some non-truth-oriented goal, presents the formation of this belief as a consequence of practical reasoning: it is an outcome of a project that is undertaken by the agent in order to fulfill a desire.8

The view according to which self-deception involves an intention to form a belief, I shall label the intentionalist account of self-deception. It constitutes the standard philosophical approach on this matter. Donald Davidson’s account of self-deception is an especially clear version of the intentionalist approach. I shall therefore appeal to Davidson’s formulations to explore this widely shared theory.9 According to Davidson, wishful thinking presents a sequence in which a desire that \( p \) produces the belief that \( p \). In contrast, self-deception requires “the agent to do something with the aim of changing his own views” (Davidson 1986, p. 87, original emphasis). Not only must the self-deceived subject bring about his own deception, he must do so intentionally:

it is not self-deception simply to do something intentionally with the consequence that one is deceived, for then a person would be self-deceived if he read and believed a false report in a newspaper. The self-deceiver must intend the “deception”. (Davidson 1986, p. 87, my emphasis)

Here is Davidson’s concise characterization:

\[ A \text{ has evidence on the basis of which he believes that } p \text{ is more apt to be true than its negation; the thought that } p, \text{ or the thought that he ought rationally to believe } p, \text{ motivates } A \text{ to act in such a way as to cause himself to believe the negation of } p \ldots \text{ All that self-deception demands of the action is that the motive originates in a belief that } p \text{ is true (or recognition that the evidence makes it more likely to be true than not), and that the action be done with} \]

8 Claiming that self-deception is thus rationalized by a set of practical reasons does not imply either that engaging in self-deception constitutes a rational action or that the belief adopted through self-deception is a rational belief.

9 The features of Davidson’s account which set it apart from other versions of the intentionalist view of self-deception, are not relevant to the present general examination of the intentionalist view. For this reason, these features are not considered in the body of this essay. The distinctive features of Davidson’s account are discussed in Lazar (1998). Davidson’s view of self-deception is presented in Davidson (1985, 1986 and 1988).
the intention of producing a belief in the negation of \( p \). (Davidson 1986, pp. 88–91)\(^\text{10}\)

The action, says Davidson, may be no more than an “intentional directing of attention” away from the evidence which supports \( p \), or may consist in an active search for evidence against \( p \).\(^\text{11}\) Thus, in an example offered by Davidson, a man believes irrationally that he is about to pass his driving test. According to Davidson, Carlos who is most averse to failure, is initially convinced, upon considering the evidence at his disposal, that he will fail his impending driving test. Carlos reasons that, other things being equal, he would rather avoid pain; therefore, Carlos would rather believe that he will pass. So, Carlos may focus on the evidence which supports his favored belief and may engage in pushing the rest of the evidence “into the background” (Davidson 1986, p. 89).

In insisting that the self-deceived person must do something in order to bring about the formation of the desired belief, Davidson, along with other proponents of this view, rightly recognizes two points. First, it is impossible to form a belief directly upon realizing that one would, all things considered, be better off if one held it.\(^\text{12}\) Second, it is recognized that the belief in self-deception corresponds to some of the subject’s practical reasons. If this view is right, then it would be impossible for Carlos to reason that

\(^{10}\)The requirement that the motivation originates from the subject’s holding the rational belief is justified only by Davidson’s endorsement of the intentionalist account. If the intentionalist account is right, then the motivation to form the irrational belief must indeed originate in the subject’s holding the rational belief. According to the intentionalist script, the subject recognizes her true predicament (and perhaps its foreseen effects) in holding the rational belief (for example, she realizes that her love is not being reciprocated and that this is going to cause much heartache). This realization causes the formation of the desire (as well as the corresponding intention) to form the negation of the presently held (rational) belief in order to avoid the unpleasant consequences associated with holding it. If, as I contend, the intentionalist account does not apply to most cases of self-deception, the requirement that the process of self-deception originates in the subject’s holding the rational belief becomes obsolete.

\(^{11}\)In his most recent article on the subject, Davidson still claims that all cases of self-deception involve an intention to form a belief. Nonetheless, towards the end of this essay, he acknowledges other forms of self-deception (Davidson 1988, p. 18). Davidson’s recent discussion can be taken to present a shift from his familiar view (most clearly presented in Davidson 1986) towards a view which no longer takes self-deception as inevitably involving an intention to form a belief. If indeed Davidson’s view of this topic changed, it is not a fully acknowledged and worked out change at this point. It is my contention, however, that even if Davidson currently endorses a more limited scope for the intentionalist account, the view is wrong in the following sense: central cases of self-deception, including those hailed by Davidson as exemplifying the intentionalist view, are rarely explainable by appeal to an intention to form a belief (for a more detailed treatment of Davidson’s account, see Lazar 1998).

\(^{12}\)For a discussion of why it is impossible to acquire a belief at will, see Williams (1973) and Bennett (1990).
believing in success is good, all things considered, and proceed to form the belief directly as a consequence of his reasoning. For this reason, the intentionalist account depicts the self-deceived subject as doing something with the intention that his action cause the formation of the desired belief.

It is clear that there are some, albeit very few, cases of self-deception that should be viewed as involving an intention to form a belief. Mostly, these rare cases involve an appeal to “cold” biasing mechanisms. Consider a person who is swayed by the force of Pascal’s wager and yet finds that he does not believe in God (Pascal 1662, pp. 233–41). He forms the intention of becoming a believer and begins to behave as one. This practice may result, after a lengthy process, in his coming to believe in the existence of God. There is a great amount of information concerning factors which facilitate formation of beliefs that do not correspond to the evidence (for example, the order in which the evidence is encountered; the context in which evidence and problem are presented). Intentional exploitation of “cold” biasing patterns in the service of self-deception is possible.

I concede at the outset, therefore, that the intentionalist account applies to some cases of self-deception. Nevertheless, only a small set of special cases of self-deception should be understood in this way. To support my claim, I present three objections to the intentionalist account: the problem of crazy choices, the problem of the negative cases and the problem of how it is done.

13 See footnote 4 above for references.
14 There is another type of case which may be explained by the intentionalist account. Consider the following example (based on Johnston, 1988, pp. 76–7). Someone is tempted to engage in a type of conduct which she finds morally reprehensible and which she expects to result in bitter self-recrimination. After engaging in this activity, she decides to take a drug that would cause her to forget, after a full night’s sleep, what she has done the previous day. Consequently, the agent does not believe that she engaged in the reprehensible activity.

This sort of case is not at the centre of discussions concerning self-deception because of the peculiar way in which it accounts for a belief’s corresponding to the agent’s practical goals. The agent appeals to mechanisms such that her knowledge of their operation and of her having set them off, does not pose a considerable impediment towards her holding the desired belief. Thus, after ingesting the drug, the subject may be aware of the general effects of the drug and of having ingested it with the intention of inducing forgetfulness of an undesirable act. This should not undermine the effect of the drug. In addition, it is not clear that, as described, this is a case of self-deception. It seems that the subject’s belief that she did not engage in the reprehensible activity (if it is indeed generated) is not formed in the presence of strong evidence to the contrary. Furthermore, even if genuine cases of self-deception may be constructed along these lines, they would be outside the scope of the present discussion. Such cases do not seem to generate the much debated puzzle of self-deception which is described at the beginning of this article.
2.2 The problem of crazy choices

The following example is based on an elegant experiment which was conducted by George Quattrone and Amos Tversky (1986). In this experiment, subjects’ tolerance to cold water changed in accordance with what they thought constituted an indication of future health and longevity. Tolerance to cold water was measured first at the early stages of the experiment. After this measurement was taken, subjects were told that each person has one of two types of cardiovascular systems: one kind is strong and the other is prone to illness. Some subjects were told that an exercise-induced increase in tolerance to cold water was correlated with a strong cardiovascular system while others were told that a decrease in tolerance indicated that happy prospect. Tolerance to cold water was measured for the second time after the subjects attended these instruction sessions and exercised for ten minutes. It was discovered that tolerance to cold water changed—increased or decreased—in the direction that in the lectures had been associated with longevity: subjects who were told that increased tolerance indicated longevity exhibited higher tolerance compared with the first measurement, whereas subjects who attended the other instruction session exhibited lower tolerance. This behavior characterized approximately 80% of the subjects. Roughly 20% percent of subjects did not exhibit a change in tolerance level. Finally, subjects were asked to fill out a questionnaire that examined whether they remembered the lecture material, and indeed the great majority of them did. They were also asked whether they purposely tried to alter the amount of time for which they kept their hand under water.

On the intentionalist account, which is endorsed by the experimenters, the subjects intentionally changed their tolerance levels to cold water so that an indication of longevity would be generated. Most subjects who shifted tolerance denied having tried to alter it, and the experimenters determine that these subjects were unaware of that plan and its consequent execution. Among the subjects who shifted tolerance, those who were

Only one subject (out of a total of thirty eight), “a suicidal type”, exhibited changed tolerance in the direction that was said to be contra-indicative of longevity.

Quattrone and Tversky argue well to the effect that the subjects’ denial is sincere. Whereas the great majority of deniers believed that they had the robust heart, most subjects who conceded that they shifted their tolerance purposely (non-deniers) concluded that they were fated to a life of illness. This, despite the fact that deniers and non-deniers where equally likely to shift their tolerance in accordance with a “positive” diagnosis (pp. 47–9). Together, these findings undermine the hypothesis that the subjects were insincere. Furthermore, subjects reported whether or not they shifted tolerance purposely by filling out an anonymous questionnaire. This procedure should weaken the subjects’ inclination falsely to deny shifting their tolerance.
unaware of a purposeful attempt to shift tolerance were more likely to think that their tolerance level to cold water was indicative of longevity, relative to subjects who were aware of such an attempt. Quattrone and Tversky contend that the subjects whose tolerance shifted, and who denied a purposeful action to this effect, were self-deceived. The self-deceived subjects, claim the experimenters, believed both that they “purposely engaged in the behavior to make a favorable diagnosis” and that they did not engage in such behavior. These subjects are said to be unaware of holding the former belief and their lack of awareness is claimed to be “motivated by [the] desire to accept the diagnosis implied by the behavior” (1986, p. 41, cf. pp. 47–8).

Two essential features of the intentionalist account are exemplified here. First, the intuitive appeal of this account is grounded in the fact that the subjects really want to enjoy a long and healthy life. Intuitively, it seems that the intensity of this desire is responsible for the peculiar behavior of these subjects. And indeed it may seem as if Quattrone and Tversky’s analysis merely appeals to this element of the subjects’ psychology in order to explain their choices. But this isn’t the case. Quattrone and Tversky’s interpretation takes the subjects’ behavior to be dominated by the goal of believing that what is in store for them is a prolonged future of good health rather than by the desire for a long and healthy future. The authors’ interpretation appeals to the desire to lead a long life in order to explain a project that is driven by the desire to believe that such a life is forthcoming. If the intuitive appeal of this analysis is to be maintained, we need to know how the former goal informs the latter.

This brings us to the second essential feature of the intentionalist account. It is indeed undeniable that many cases of self-deception involve a goal of the subject’s such that the goal would be satisfied if the irrational belief were held (for example, the goal of feeling less anxious). If the subject recognizes this fact, it may be claimed, the subject then has a desire to form the irrational belief. Under such a liberal view of desire attribution, it may well be true that self-deceived persons desire to form the irrational belief. It should be stressed, however, that the intentionalist account makes a considerably stronger requirement in this context: not only must the self-deceived subject have a desire to form a belief, but she must act on it. A serious problem for this account arises here since the goal of holding the desired belief often conflicts with the goal of making it true.

The intentionalist account requires that the subject have a desire such as “that I believe that I will pass the test”. The subject’s assessment that the belief in question is false need not be part of the content of the desire to believe. A similar point applies to the intention to form a belief. However, the intentionalist account views the self-deceived subject as being motivated by her conviction that the desired belief is either false or at least undermined by the weight of the evidence.
Returning to the experiment it is far from obvious that all of the subjects who shifted their tolerance preferred to believe in a rosy future over receiving a genuine indication of what was in store for them. In fact, it is reasonable to suppose that many subjects, if given the option, would prefer knowledge over short-lived comfort. This seems even more compelling in a case, such as is presented in the experiment, in which such knowledge may directly affect the length of one’s life. Prior knowledge of a vulnerable cardiovascular system may prolong one’s life by leading one to undertake proper preventive measures (for example, following a proper diet, exercising and having one’s physical condition professionally monitored). Furthermore, such knowledge may produce emotional preparation which will greatly improve the quality of one’s life when illness strikes.

The intentionalist will have us believe that the desire for longevity is so strong that it causes the subject to choose to believe, against the evidence, that its fulfillment is forthcoming. So strong is this desire, according to this account, that the belief in its fulfillment is chosen over an attempt to bring about fulfillment or to prepare for frustration. The problem of crazy choices undermines the central prima facie advantage of the intentionalist account, the explanation of irrationality by appeal to reason. According to this view, the self-deceptive state is formed as a consequence of practical reasoning—of the subject’s choosing to believe what is, according to her, undermined by the evidence. Specifically, the formation of the irrational belief is taken to stem from the subject’s choosing to advance a non-truth-oriented goal (or goals). The intentionalist explains the failure of theoretical reason in self-deception by showing this failure to be the consequence of the workings of practical reason. The intentionalist account seems appealing so long as the choice of the self-deceived person is reasonable. If this choice is crazy, the intentionalist strategy fails: the intentionalist ends up explaining the breakdown of theoretical reason by appeal to the breakdown of practical reason. No explanatory advantage is gained by this shift.

2.3 The problem of the negative cases

Some cases of self-deception do not warrant the attribution of a desire to believe. Consider a case in which self-deception results in the formation of a belief more negative than is warranted by the evidence. The thought that his wife might be having an affair has crossed John’s mind in the past. When examining the evidence, John always concluded that it strongly supported the view that she was faithful. Last week, however, when his wife called to say she would be arriving late from work, John was struck by the thought that she was having an affair. He became convinced of it and, painfully, is now obsessed with his wife’s imaginary infidelity. A calm consideration would reveal that John has no substantive evidence to
support his view. In fact, his wife has always been very affectionate, is keen on spending time with him, etc. John, however, is sure that her behavior is intended to mislead him. He constantly searches for “incriminating” evidence, convinced that, sooner or later, he is bound to discover it. John is already convinced of his wife’s infidelity; he desperately wishes to attain strong evidence so that he could make her admit that she is having an affair. Negative cases such as this present a problem for the intentionalist: they seem to undermine the claim that the formation of the irrational belief in self-deception is always due to a desire to form it.

The intentionalist can take one of two routes in defending her position with respect to such cases: she can insist that the irrational belief is due to a desire to believe or she can deny that such cases exemplify self-deception. Taking the first route, the intentionalist will claim that John’s irrational belief in his wife’s infidelity results from his desire to so believe. John may desire to believe in his wife’s infidelity in order to terminate a painful obsession due to uncertainty: he prefers to believe the worst over prolonging a state of uncertainty. Or John may wish to terminate the marriage and knows that he is more likely to muster the courage to do so if he believed that his wife were having an affair. One need not deny such possibilities in order to see that the suggested strategy cannot be universally applied to negative cases in a convincing manner. Most cases of irrational jealousy, for example, preclude these and similar explanations that attribute to the subject of self-deception a predominant desire to form the irrational belief. Thus, many irrationally jealous spouses fail to exhibit extreme aversion to uncertainty or a desire to break up the marriage.

The second route that may be taken by the intentionalist in an attempt to fend off the threat of the negative cases is to claim that all negative cases of irrational belief formation fall outside the scope of self-deception. This contention could be grounded in two different ways. First, the intentionalist may claim that, in contrast to genuine cases of self-deception, the irrationality in the negative cases is never due to the evaluative attitudes of the self-deceived subject. Rather, it will be said, the irrationality is due to the subject’s intellectual incompetence (to detect the irrationality of her own belief) or is a result of a different kind of mistake that is not attributable to any of her evaluative attitudes. This route does not lead to an effective defense of the intentionalist position either, since it does not accord

---

18 There are two kinds of seemingly intentional cases in this area. In the example above, John only desires to believe that his wife is unfaithful, but does not desire that she be unfaithful. In addition, it is easy to construct examples of seeming “negative” cases where the agent also desires that the state of affairs represented by the desired belief obtain: John may also desire that his wife be unfaithful (as well as desiring to believe that she is) if he strongly wishes to feel morally superior to her. Even though this belief will cause a lot of pain, it will also bring satisfaction associated with a sense of superiority.
with the phenomena. Typically, negative self-deception is exhibited around areas of particular importance in a person’s life: areas that are associated with strong desires and emotions. Indeed, many cases of irrational jealousy are due precisely to a person’s strong desire that his love be reciprocated. Familiar tendencies of underrating one’s performance are often characteristic of people who are strongly motivated to succeed in the relevant type of endeavor.

But we need not rely solely on the interpretation of isolated cases. We may appeal to well-documented patterns of irrational belief formation which stress the causal connection between desire and (negative) irrational belief. In a wonderful experiment, high school and college female students were asked to pair themselves with a classmate whom they knew best. Both members of a pair were asked to compare their appearance to each other. These subjects consistently rated their appearance as inferior to their partner’s vis-à-vis the following features: nose, hair, eyes, smile, attractiveness, complexion, figure, and expensive clothes (Koehler et al., unpublished). In all other dimensions which these students were asked to consider, personality, relationships and academic quality, they overrated themselves: most judged they fared better than their partner on a wide array of traits in these dimensions. In a corresponding male population of students, subjects consistently overrated themselves on all four dimensions (including appearance). Two points are nicely illustrated by this result. First, it is clear that a good portion of the female students’ “negative” judgments concerning appearance are irrational. Second, the “negative” irrational beliefs within the female student population are thematic: they are due, in part, to the great desire to be attractive, which characterizes these subjects.

Another attempt to ground the claim that the negative cases fall outside the scope of self-deception may consist in claiming simply that, in everyday discourse, the term “self-deception” applies only to cases which result in a more positive belief than is warranted by the evidence. In response, it should be said that even if the negative cases do not exemplify self-decep-

---

19 Some of the “negative” comparative judgments must exemplify an underrating since both members of a pair cannot be less attractive than the other. Furthermore, it is reasonable to suppose that, in many instances where the judgment is false, the evidence at the subject’s disposal undermines it. Hence, many of the “negative” judgments are irrational.

20 Examples of this kind abound. Consider the ambitious young executive who tends to undermine her own performance while judging accurately the performance of others. Her consistent underrating of her own performance is due precisely to her strong desire to excel. Similarly, a young lover who has not confessed his love may be convinced that his beloved does not care for him by interpreting every sign of impatience or lack of interest on her part as indicating this sorry state of affairs. A calm observer would conclude, on the other hand, that the woman in question is e.g., being preoccupied with another matter or lacking in self-confidence. This phenomenon will further be explored in §3 below.
tion, they are effective in undermining the intentionalist view. The similarity between the positive and the negative cases is striking—both types exemplify the formation of belief that is undermined by the evidence and is thematically linked with the subject’s evaluative attitudes. Unless a strong case is made on its behalf, there is considerable disadvantage to a theory that takes the positive and the negative cases as exemplifying completely different psychological phenomena. The intentionalists do not provide an argument for such a perspective and it is difficult to see what argument might be devised in its defense.

2.4 The problem of how it is done

The final objection to the intentionalist account relates to the problematic nature of a project that is undertaken with the intention of forming a belief. Returning to Davidson’s example, he suggests that it is most plausible to explain Carlos’ self-deception by taking him to be doing something (for example, concentrating on favorable evidence such as his charm and the effect that it has on others) with the intention of acquiring the belief that he will pass his driving test. Let us assume that Carlos is thinking that his charm has gotten him out of trouble in the past and that there is some slight chance that it will get him out of trouble in this case. Yet, if Carlos knows why he is focusing on the favorable evidence, and if he believes that there is ample evidence which supports the negation of this belief, the presence of the favorable portion of the evidence may not have the force required to induce the positive belief. Thus, even if Carlos’s concentration on the favorable evidence makes it, for a moment, seem more significant than it truly is, the realization that it is not all that significant, but only appears to be so due to a manipulation, is likely to inhibit the formation of the desired belief. Rather than explain the formation of the irrational belief, the presence of an intention or plan to form a belief constitutes an impediment towards its formation since it is likely to interfere with the effect of biasing mechanisms. For this reason, explanations of the formation of the irrational belief that do not appeal to an intention of this kind have a prima facie advantage over the intentionalist account. I shall suggest a non-intentionalist account in the following sections.

The intentionalist may not concede this point so easily. She may claim that the intention to form the irrational belief does not stand in the way of forming the belief as long as the subject is not aware of it. In response, I

21 In a similar move, the intentionalist may claim that if the agent is unaware that she is engaged in an activity which is intended to bring about the formation of the desired belief, the project may be completed smoothly. In the body of the article, I deal only with the intentionalist’s objection in terms of the agent’s unawareness of her intention to form a belief (rather than her being unaware of the nature of her action). Both versions of the objection are essentially identical, hence my response to the first version may be applied to the second version with obvious changes.
shall argue that this objection cannot salvage the intentionalist account since the intentionalist cannot answer the question of why it is that the agent is not aware of her intention to form the irrational belief.

There are two more or less recognized basic options for explaining the agent’s lack of awareness. Either the correspondence—between an agent’s unawareness of the intention to form a belief, on the one hand, and the satisfaction of a desire, on the other—is to be given a non-thematic explanation, or else it is to be explained by the agent’s intentionally “keeping it away” from consciousness as part of a plan which is geared towards goal attainment.

In the non-thematic type of explanation, the content of the intention to form the belief and, particularly, its relation to goal attainment play no role in explaining the agent’s unawareness of the intention. Such cases are familiar. Take, for example, my lack of awareness of intending to pick up the apple in front of me while engaged in a fascinating conversation. This lack of awareness, the intentionalist will rightly say, is not due to any special process or action on my part; nor is it related to the specific content of the intention. Rather, such unawareness ought to be explained by general facts and principles pertaining to information processing. While the study of such facts may be of great value in the context of exploring the nature of cognitive processes, it will be claimed that it is not especially relevant to self-deception. The intentionalist will emphasize that unawareness of mental states is a quite general phenomenon. She would therefore deny that she carries a special burden in explaining it as part of her account of self-deception.

In light of this argument, the intentionalist’s position is as follows. The agent’s lack of awareness of the intention (to form a belief) explains why it resulted in a successful execution (why the agent acquired the desired belief). Nevertheless, this option determines that the agent’s lack of awareness was not due to its instrumental value. It is simply a “lucky draw”, from the point of view of the planning agent, that due to other factors, she never became aware of this plan. In other words, the intentionalist acknowledges that self-deception is thematic—that the formation of the irrational belief is explained by appeal to its content. This is why she appeals to practical reasoning in the first place (for the full argument, see §§1 and 2.1 above). At the same time, lack of awareness of the intention to form the belief is explained non-thematically that is, by appeal to general facts about cognition which are by and large content independent.

There are a few problems here. First, if the self-deceived agent’s lack of awareness is on a par with my lack of awareness of intending to pick up the apple, then we should expect that people would be regularly aware of their intentions to form discredited beliefs. If I am not fully engaged in the conversation or am simply paying more attention to my thoughts, I may realize
that I intend to pick up the apple. But one is never aware of an intention to form a discredited belief, even though self-deception is a rather common phenomenon. Even if one’s attention is drawn to the relevant subject matter, one does not simply become aware that one is planning to get oneself to believe that one’s love is being reciprocated. In contrast, one may become aware that one holds a certain belief, or that one wants something to be the case. Second, the intention to form a belief corresponds to a very complex action (unlike the action of picking up the apple). Intentions of this kind, in contrast with simple intentions to pick up an apple, are less likely to be unconscious for non-thematic reasons. Third, if the intentionalist is right, people who realize that they intend to form a belief in self-deception would have their plans shattered right away. They would think: “What bad luck, now that I found out about my plan to form the belief, I will never believe that Maggie is faithful!”. Put mildly, this is not a familiar phenomenon. Finally, one would be hard pressed to identify the general principles and facts which explain unawareness of the self-deceptive intention.

The second candidate for explaining the agent’s unawareness of the hypothesized intention to form the irrational belief is better in one respect: it recognizes that the agent’s unawareness of the intention is related to its content. According to this solution, the agent intentionally keeps the intention away from consciousness (either by doing something which will prevent her from becoming conscious of the intention in the first place, or by doing something which will make it the case that she is no longer aware of it). This solution reiterates the intentionalist account of self-deception. This time around, the agent’s unawareness of the intention is itself a consequence of an intention. But this solution merely pushes the problem of self-deception one step further. Either this second-order intention is itself unconscious (in which case we want an explanation of this fact); or else it is conscious and the project seems self-defeating.

This does not apply to very rare cases which are explained by appeal to an intention (see §2.1 above). Even these cases fail to meet the condition described directly below.

This strategy of defending the intentionalist view runs into another problem. Why, it may be asked, does the unconscious intention fail to exhibit its normal effects? Why does it not impede the formation of the desired belief? After all, unconscious states produce many of their standard conscious effects (for example, guilt, anxiety, derivative beliefs). It is precisely these effects which justify our attributing unconscious states to agents in the first place.

My criticism of the intentionalist strategy is not intended to diminish the role of unconscious thoughts in the explanation of behaviour and it is certainly not intended to show that thematic attribution of unconscious thought is self-defeating. Rather, it shows that thematic attribution cannot be understood along the lines of “putting away” thoughts from consciousness. My criticism is merely intended to establish a negative point against the intentionalist. A positive account of the presence and function of unconscious thoughts is well beyond the scope of this article.
The three arguments presented so far discredit the standard approach to self-deception. Since “cold” biasing mechanisms also fail to account for self-deception, we must look for a new approach. In the following section, the systematic effects of desires and emotions on the formation of beliefs are presented as the basis for a new account of self-deception.

3. Broadening the perspective: emotions, desires and cognition

When I am angry, what my best friend did that upset me looms large, and her kindness towards me seems rather insignificant. I may be aware of being angry as well as of the common cognitive effects of this emotion. Yet I may still fail to see her in a more favorable light. I take her to be an inconsiderate, self-centered person.

The arousal of emotions is of great relevance to belief-formation. Emotions affect information processing skills of various kinds: they are associated with less systematic thinking, less efficient processing skills, reliance on simplistic response strategies, decreased reliance on direct evidence, and increased reliance on superficial cues. In addition, emotions affect attention and salience patterns, as well as working memory capacity.24 There is evidence that belief formation under the influence of an emotion is strongly affected by selective features of the environment and is removed from systematic reasoning. This is so because, in addition to causing considerable reliance on superficial cues, emotions affect the salience level of environmental features: what is more easily observable and stands out—in one’s environment as well as in one’s memory—is highly sensitive to the subject’s emotional state. Furthermore, emotional arousal causes superficial consideration of the evidence. It is all the more easy to form a distorted view of the evidence in the absence of careful consideration.

A distorted assessment of the evidence may be maintained even if, in the past, one has considered the evidence carefully. A person may be aware that, in the past, he considered his friend as most reliable, yet in the grip of envy or anger, the friend appears less virtuous. The earlier view of one’s friend, even if it is considered, may be easily discarded as mistaken or as having

been poorly recorded in one’s memory. Even when one’s attention is directed to past experiences when one’s friend has been helpful and loyal, the negative aspects in the friend’s personality still loom large.

The role of emotions in the formation of irrational beliefs has been largely ignored in discussions of self-deception. Overwhelmingly, the formulation of the problem of self-deception does not appeal to emotions at all, and most often, neither does the solution. Yet, convincing cases of self-deception involve beliefs which are associated with very strong emotions in the subject. I cannot think of one clear example of self-deception where the irrational belief (as well as its rational counterpart) is not associated with a strong emotional response. Indeed, the intensity of an emotion may be such that, in an extreme case, a person may hold a view while admitting that it is undermined by the weight of the evidence. A case in point is the denial of a loss of a loved one in the face of recognized overwhelming evidence to the contrary. When notified of her son’s death in battle, the distressed mother rejects this proposition despite conclusive evidence for the tragic state of affairs. She may say that, “in her heart”, she knows that her son is alive. Indeed, the lack of correspondence between belief and evidence while the subject is in the grip of an extremely intense emotion is very familiar. It is often described in everyday discourse as a “refusal to believe”.

The significant effect of emotions on belief formation is exhibited in positive as well as negative cases of self-deception. More importantly, and this is the heart of this point, cognitive shifts that are produced by emotions, occur in the absence of changes in relevant information: the subject “puts together” the evidence in a different way when her mood shifts. Recall John who, in the grip of intense jealousy, sees “incriminating” evidence concerning his wife’s behavior wherever he turns. In the grip of jealousy or rage, every aspect of his wife’s behavior seems suspicious, while her affectionate behavior and consistent support are not given their due weight. Often, the self-deceived jealous person experiences periods in which he is convinced or even sure that the evidence of betrayal is overwhelming, as well as periods when he realizes that he does not possess such evidence. These cognitive shifts need not be accompanied by changes in relevant information. Or consider the following common case. Mark has just lost his job and realizes that it will be difficult to find new employment. He is anxious and his future appears bleak. After spending a few hours in cheerful company, Mark is convinced that he can find a job and support his family. Mark need not have acquired new evidence within this period of time in order to enjoy this new outlook on life. Rather, in merry company, joined by his family and
friends who support him, Mark’s mood changes considerably and the future suddenly appears rosy.

The phenomenon of depression provides many vivid examples of radical changes in belief in the face of the same body of evidence. Depressed persons undermine their prospects for success, their skills as well as their performance.\(^\text{25}\) In a wonderful illustration, Peter Kramer (1993) describes an adult depressed patient who has always felt inferior to her cousin whom she knew well. The cousin is not only experienced as superior in her achievements and talents, but also in her appearance. In particular, the patient views herself as short and chubby, the cousin as lean and tall. Once on an anti-depressant, the patient discovers for the first time that her cousin is really shorter than she is.\(^\text{26}\)

One factor which contributes to the baffling nature of the traditional puzzles of self-deception is ignoring the effect of emotions on thought. In considering someone else’s self-deception calmly, we fail to understand how a person may come to hold the irrational belief, given the undermining evidence at her disposal. Even when we consider that strong desires are involved, we still fail to see how the irrational belief may be formed. After all, we reason, there are a number of things we desire, yet we (often, painfully) distinguish between desire and its fulfillment. Our bafflement about the self-deceptive belief will be reduced if we realize that in the grip of a strong emotion, the world may appear very different. The assignment of a central role to emotions in the formation of self-deceptive beliefs is largely incompatible with the view of self-deception as an action. Emotions do not affect one’s view of the world through deliberation: they do so immediately and in a way which, to a high degree, is not subject to our control. Fear, hope, and their kin (and in concert with other elements) affect cognition without the existence of a plan that is designed for this purpose.

There is a similar case to be made concerning the effects of desires on cognition. A case in point is the experiment concerning self-evaluation versus other-evaluation mentioned above (Koehler et al., unpublished). We saw (in §2) that, in contrast to other areas where female students overrated themselves, their own personal appearance was consistently underrated. This phenomenon seems related to the girls’ strong desire to be attractive. Other effects of desires on cognition are also highly predictable and do not exhibit sensitivity to factors which would be at play were the subjects attempting to form a self-serving judgment intentionally. A case

\(^{25}\) See e.g. Rosenfarb and Aron (1992), Slife (1992) and Simons et al. (1993).

\(^{26}\) This example illustrates nicely the effects of emotion on perception which have not been emphasized in the text. It was pointed out to my by Nomi Arpaly. Further material is found in Niedenthal and Kitayama (1994).
Deceiving Oneself Or Self-Deceived? 283

in point is the manipulation of a belief by frustrating an unrelated desire (Dunning et al. 1995).

Subjects who were interested in pursuing a post-graduate degree failed an intellectual task which, they were told, was highly correlated with a successful post-graduate career. A control group of subjects were given an easier task and passed. Those subjects who failed were much more inclined to make self-aggrandizing judgments concerning their likelihood of enjoying a successful marriage. All subjects were presented with a selection of personal profiles. Those in the failure condition judged that the persons who most resembled them were most likely to enjoy a successful marriage. In addition, when presented with a profile of a person who, subjects were told, is enjoying a good relationship with their spouse, those in the failure condition selected the features which they shared with that fortunate person (for example, being the oldest sibling) as explanatory of the latter’s successful marriage. In contrast, subjects in the success condition (those who were made to pass an initial test) did not exhibit these biases. Examples such as this are telling, since the difference between the failure and success conditions concerns skills entirely different from those required for a successful marriage. Such examples cannot be accommodated within the intentionalist solution since we must suppose that subjects in both groups were equally qualified and equally motivated to enjoy a good marriage. We see that the effect of a frustrated desire (for success in graduate school) on cognition (concerning an entirely different subject-matter) has the appearance of a “blind”, non-rational causal process.27

One lesson to be drawn from the array of phenomena sampled above is that motivational and emotional factors are regular players in determining what beliefs we hold. We should also accept that the philosophical puzzle concerning the process by which the self-deceptive belief is formed, is only produced if we ignore the large role of these factors in the formation of beliefs. Appeal to practical reasoning in order to explain these effects is hopeless in light of their prevalence and variety. In contrast to the predictions of the intentionalist approach, the formation of emotion-induced (or desire-induced) beliefs is highly uniform, quick and effortless.

In light of these phenomena, we also lose the hope of uncovering one simple structure that underlies all those cases which we label self-deception. To the contrary, we are led to recognize a variety of elements which play a role in determining belief acquisition as well as a variety of patterns by which beliefs are formed. A detailed analysis of instances of self-

27 Examples of other work which explores this phenomenon include Kunda (1987), Kunda and Sanitioso (1989) and Sanitioso et al. (1990).
deception must be made with attention to the particulars of each case and in conjunction with experimental findings.28

4. Fantasies and the imagination

There are a number of ways in which imagination is relevant to self-deception. One familiar way in which what we imagine affects what we believe is through triggering emotional responses. Often, one imagines how one would feel, what one would do, how one would respond, if a certain state of affairs were to obtain. It is not uncommon that emotions get aroused in this process and, in turn, affect one’s view of the evidence. A second way in which what we imagine may affect our beliefs need not involve arousal of emotions. Through imagining a situation in detail, the likelihood of its obtaining may seem greater than is suggested by the evidence. Someone who enters an important athletic competition may imagine herself dominating it. The athlete imagines how she would perform perfectly at the crucial time, reviewing in her mind the different tasks that she would be required to undertake. She proceeds to imagine the judges’ positive assessment of her performance and the favorable outcome. Executing (in her mind) all the required tasks to the point of perfection may convince our competitor that an excellent performance is well within her reach even in the presence of undermining beliefs. This phenomenon is widely recognized by athletic trainers who instruct their trainees to engage in such exercises as a way of increasing self-confidence.29

Imagining is revealed in these two kinds of cases as a process through which emotions and desires affect cognition. But imagination may also aid us in providing a conceptual framework for describing the end product of the process of self-deception. For this purpose, we ought to distinguish between two different ways in which people imagine a frustrated desire being fulfilled. Richard Wollheim points out that there are two different types of activities which fit this description (1984, pp. 84–92). One type

28 Cognitive and motivational factors also affect the perception of pain. A more convincing alternative interpretation of the results of Quattrone and Tversky’s experiment discussed in §2.2 above is that the subjects’ desires concerning tolerance for cold water directly affected their beliefs about their tolerance levels. Such beliefs, in turn, may affect pain thresholds. Interestingly, Quattrone and Tversky (1986) briefly consider the possibility that the subjects’ motivation may affect pain threshold directly and suggest that it is worthy of further consideration. This alternative, however, is not reflected in their final analysis of the case.

29 Most of the work in this area explores the connection between imagery and performance levels. Some of this work examines the effects of imagery on cognition. For relevant material, see Murphy (1990), Smith (1991) and Suinn (1985).
is properly viewed as an attempt to discover the various consequences should one’s currently frustrated desire be satisfied. John Stuart Mill’s famous exploration of the consequences of having all of his desires satisfied is an excellent example of this kind of activity (1874). In this type of case, what is imagined reflects many of one’s relevant beliefs and desires. Failure to incorporate these elements in the imagined scenario undermines the project and the outcome of such an exploration may not be taken as a reliable indication of what would occur, should the desire in question be fulfilled. This type of activity is essential to practical reasoning.

Imagination, however, functions not only as a vehicle for exploring ways of satisfying a desire and the consequences of having it satisfied, but it is also a means through which desires are expressed. It is activity of this kind which is especially relevant to the discussion of self-deception. After attending the great ball at La Vaubyessard, Emma Bovary imagines that she is invited back to the château as a special guest, frequents society events in Paris on a regular basis and, attired in the most exquisite gowns, converses with ambassadors and dukes (Flaubert 1857). These dreams do not represent Madame Bovary’s knowledge of her melancholic dispositions, her dissatisfaction with herself, her financial situation, and her inappropriate social background for such sustained interactions. Her dreams do not reflect an attempt to find out what would most likely happen if she were invited back to the château. Rather, they are expressive of her desires for attention, adoration, etc.

As expressions of desires, fantasies of this sort do not reflect one’s beliefs concerning, for example, the social consequences of one’s fulfilling the desire, the full range of one’s feelings concerning such fulfillment, or the relevant features of one’s personality. Despite the fact that Madame Bovary’s dreams are incompatible with some of her own beliefs, she is convinced that, if given the slightest opportunity, she could become a thoroughly satisfied star of Parisian society. In other words, Madame Bovary does not conceive of these states (for example, the conviction that she is likely to become a contented star of Paris society if provided with the slightest opportunity) as mere figments of her imagination, but rather takes them to be representative of reality. Much like the athlete mentioned above, the more Madame Bovary allows herself to live in fantasy, the more convinced she becomes that she has uncovered the solution to all that ails her.

In referring to these states of Madame Bovary’s as “fantasies” or “dreams”, I intend to emphasize the extent to which they are detached from evidence which Madame Bovary herself possesses. I am not thereby denying that these are instances of believing. In my view, beliefs range enormously from “model” cases to crazy constructions which obviously clash with the bulk of the subject’s beliefs.
In one respect, the self-deceptive state resembles simple cases of fantasy of the daydreaming kind. Just like a daydream, this state expresses the subject’s fears, hopes and wishes. Like a daydream, it conflicts with the bulk of the subject’s beliefs. However, whereas a simple daydream exerts a rather restricted influence on one’s choices and thinking, the self-deceptive state affects one’s behavior no differently than most beliefs: it figures in both practical and theoretical reasoning. Emma Bovary spends a large amount of money, indeed brings her family to financial ruin, because she is convinced that the appropriate outfits will get her the coveted opportunity to enter society or to attract a suitor of high social standing. While this conviction influences Madame Bovary’s actions via reasoning (for example, she reasons how to get the money, how best to spend it), its content fails to reflect crucial information. It merely represents her wishes for being grand, adored, unique.

States of self-deception are a sort of hybrid: they are strongly influenced by desires and emotions—they express them—yet they inform our behavior more or less in the way that beliefs do. We need not be puzzled on encountering a state which functions much like a belief does even though its content primarily corresponds to a desire. A great number of beliefs deviate from norms of rational belief formation. Instances of belief formation range from extremely well justified cases through beliefs which are significantly colored by desires and emotions. Still further on this spectrum, we find beliefs which are considered “mad” for being so far removed from standards of rationality. We occasionally run into people who believe that Martians sent them to earth or that most security services monitor their activities. Self-deception is best understood within this wide range. Qua irrational beliefs, self-deceptive states are not unique. They make us feel uncomfortable because states of self-deception threaten the conviction that beliefs and evaluative attitudes “don’t mix”: that they are generated through entirely different kinds of processes. Ideally, beliefs ought to be formed exclusively in response to epistemically relevant factors. The psychology of the formation of beliefs, however, suggests that they are systematically sensitive to a wide range of factors, many of which are epistemically irrelevant: factors which do not constitute evidence either for or against the belief. In particular, beliefs are sensitive to motivational factors. But beliefs are not merely caused (in part) by desires and emotions. More interestingly, beliefs may express desires when their contents reflect, to a higher or lower degree, what is desired, feared or hoped. Thus, a lover’s irrational conviction that he will be rejected may express his strong desire that he not be rejected (and perhaps his fear of being rejected) more than it mirrors the evidence at his disposal. In a similar fashion, beliefs may
express one’s moods and emotions rather than strictly reflect what is epistemically warranted.

I have argued that very few cases of those labeled self-deception may be understood by taking the agent to be deceiving herself. Rather, the self-deceptive belief is caused, in part, by desires and emotions whose effects are not mediated by practical reasoning. Both the object of desire and the object of fear may be reflected in the contents of our beliefs (for example, one may believe, contrary to the evidence, that one’s love is reciprocated, or one may believe, contrary to the evidence, that one’s love is not reciprocated). We label self-deception those beliefs which are at odds with the evidence and which are highly influenced by the presence of emotions and evaluative attitudes. The unintentional, more direct, effect of wishes has been mentioned in the philosophical literature and is usually termed “wishful thinking”. I have attempted to show that wishful thinking—and fearful thinking—is a much more central phenomenon in the process of belief formation than is currently thought. I have also argued that the most extreme and difficult cases of motivated irrational belief formation are better understood along these lines.

The array of phenomena surveyed in this essay provides but a glimpse of the enormous influence which evaluative attitudes, moods and emotions exert on the formation of beliefs. Such effects are by no means exclusively exhibited in the dramatic cases we distinguish by special labels such as “self-deception”, “wishful thinking” or “crazy belief”. Rather, in a less dramatic fashion, what we want, hope or fear, is often expressed in the way things seem to us.\(^\text{31}\)

\(^\text{31}\) I received valuable comments from Michael Bratman, Jean-Pierre Dupuy, Michael Della Rocca, Dagfin Føllesdal, Edward Harcourt, Maria Merritt, John Perry, Lee Ross, an anonymous referee and, especially, Donald Davidson and Richard Wollheim. Many thanks are due to Bernard Williams for extended conversations on this topic.

References


**Department of Philosophy**

*Northwestern University*

1818 Hinman Avenue

*Evanston, IL USA*

a-lazar1@nwu.edu

**ARIELA LAZAR**

**REFERENCES**


