Croatian Journal of Philosophy
Vol. XIV, No. 42, 2014

From Self-Deception to Self-Control: Emotional Biases and the Virtues of Precommitment

VASCO CORREIA
Insituto de Filosofia da Linguagem
Universidade Nova De Lisboa

'Intentionalist' approaches portray self-deceivers as "akratic believers", subjects who deliberately choose to believe p despite knowing that p is false. In this paper, I argue that the intentionalist model leads to a series of paradoxes that seem to undermine it. I show that these paradoxes can nevertheless be overcome if we accept the hypothesis that self-deception is a non-intentional process that stems from the influence of emotions on judgment. Furthermore, I propose a motivational interpretation of the phenomenon of 'hyperbolic discounting bias', highlighting the role of emotional biases in akratic behavior. Finally, I argue that we are not the helpless victims of our irrational attitudes, insofar as we have the ability—and arguably the epistemic obligation—to counteract motivational biases.

Keywords: Akrasia, emotions, epistemic responsibility, hyperbolic discounting, irrationality, motivational biases, precommitment, self-deception, self-control.

1. *Introduction*

Audi suggests that "a philosophy of mind that cannot account for [self-deception] is seriously deficient, and a psychology that says nothing about it is unwarrantedly narrow" (1988: 92). Yet, self-deception poses both a descriptive and a normative challenge. From a descriptive point of view, the difficulty concerns the very possibility of self-deception, particularly if we accept the influential 'intentionalist' model (Davidson 1985a, Pears 1984, Sartre 1969, Scott-Kakures 1996, Gardner 1993, Bermudez 1997), which maintains that self-deceivers typically get themselves to believe that *p* is true, knowing all the while that *p* is

false. From a normative point of view, on the other hand, there is considerable controversy over two independent questions: First, whether self-deception may or may not contribute to people's overall happiness (for a review, see McKay & Dennett 2009); and second, whether it is morally acceptable to deceive oneself (Barnes 1997, Clifford 1994, Martin 2009).

In this paper, I initially address the descriptive question. I critically examine the postulates of the intentionalist model, and argue that it is inextricably paradoxical, leading inevitably to at least one of four paradoxes (sections 2 and 3). In section 4, I argue that the rival 'motivational' account proposed by Mele (1987, 2001a) succeeds in overcoming each of these paradoxes. Further, I show that this view is consistent with the empirical studies carried out by social psychologists upon the topic of judgment biases. According to it, self-deception and other phenomena of motivated irrationality (wishful thinking, rationalization, motivational biases, etc.) stem from the influence that strong emotions exert over the process of belief formation. But emotional influences are also liable to affect practical judgments and the process of decision making. In section 5, I bring forward a motivational version of the phenomenon of 'hyperbolic discounting' that seems to account for ordinary cases of akrasia (Ainslie 2001, Elster 2007, Loewenstein et al. 2003). In light of this hypothesis, it is the influence of an affective state (emotion, desire) that triggers the temporary preference reversal thought to be the root of akratic action. Finally, I argue that we have the ability—and perhaps the moral obligation—to counteract the effects of motivational biases upon our judgments (epistemic self-control), either by controlling the process of belief formation or by resorting to indirect strategies of precommitment.

2. *Deciding to believe*

According to the intentionalist account (Davidson 1985a, Pears 1984, Sartre 1969, Scott-Kakures 1996, Gardner 1993, Bermudez 1997) self-deception is not an accidental phenomenon that happens to the agent in spite of him, but rather an intentional process whereby the self-deceiver adopts a false belief that seems to enhance psychological comfort. This is allegedly what happens, for example, when a terminally ill patient persists in believing that she will survive despite the unanimous prognosis of several doctors. Davidson evokes the similar example of Carlos, an individual who knows he will probably fail the test for a driving license, given his lack of preparedness, but manages to convince himself that he will succeed, despite being "aware that the totality of the evidence points to failure" (Davidson 1985a: 207).

To be sure, this kind of attitude is clearly irrational from an *epistemic* point of view, given that the desire to believe that p is surely not a good reason to believe that p. However, from a *practical* (or utilitarian) point of view, it is arguably rational to prefer a sweet illusion to

the bitter truth. In fact, the criterion of practical rationality generally adopted by decision theorists is a purely formal one which is indifferent to the question of truth: The main requirement for an action to be rational is that it remains consistent with the agent's preferences; in other words, that it seeks to maximize the agent's well-being. And from that perspective, both Carlos and the terminally ill patient seem to have legitimate reasons to believe (what they suspect to be) a falsehood. Carlos's unrealistic optimism is justified by the fact that he cannot bear the thought of failing the test again, let alone the possibility that his optimism might bring him some extra confidence and actually increase his chances of passing the test successfully. As to the terminally ill patient, similarly, there seem to be valid practical reasons to adopt a positive illusion: First, it will spare her the anxiety of thinking that death is very near; second, it will bolster her motivation to fight the disease (which might contribute to prolong her survival); and third, it will make her happier during the little time she has left to live.

To this extent, the decision to deceive one-self is very similar to whatever practical decision, although its outcome is a belief rather than an action. As Davidson (1985a: 207) suggests, the self-deceiver's "practical reasoning is straightforward": He begins by deliberating that it is preferable, all things considered, to adopt the false belief that *p*; then forms the intention to deceive herself; and finally, he acts accordingly, for the process "requires the agent to do something with the aim of changing his own views" (ibid.). Hence, self-deception can be described as an *action* in its own right, much like other-deception, and some philosophers even suggest that self-deception is a sort of akratic behavior in the cognitive sphere (Heil 1984, Rorty 1983), while others go as far as to call it an act of "epistemic cowardice" in the face of anxiety (Johnston 1988: 85, Barnes 1997: 172).

But if self-deception is an action, what sort of action is it? Most intentionalists acknowledge that people are not at liberty to believe that *p* merely because they wish to believe that *p*[1]. On the other hand, it seems possible to cause oneself to believe something via some *indirect* strategy. Intentionalists mention several strategies of self-induced deception, some more plausible than others. First, the self-deceiver may try to simulate a belief, *acting as if* it were true, with the hope that it might eventually become a sincere conviction. This is the method Pascal (1958: 233) famously recommended to those who wished to persuade themselves of God's existence: "Follow the way by which they began: that is by doing everything as if they believed, by taking holy water, by having Masses said, etc. Naturally, even this will make you believe and will dull you". Second, one may resort to suggestion techniques, such as hypnosis (Williams 1973) or even self-hypnosis (Naylor 1985). Third, it is sometimes possible to manipulate the available information in such

[1] For a fuller analysis of the paradoxes of *direct doxastic voluntarism*, see for example Bennett (1990) and Williams (1973).

a way that the desired belief will appear more plausible in the future. McLaughlin evokes a stratagem of this kind:

> In order to miss an unpleasant meeting three month ahead, Mary deliberately writes the wrong date for the meeting in her appointment book, a date later than the actual date of the meeting. She does this so that three months later when she consults the book, she will come mistakenly to believe the meeting is on that date and, as a result, miss the meeting". (McLaughlin 1988: 31)[2]

And finally, the self-deceiver may also try to control the direction of his attention, focusing on the evidence that seems to support the belief she wishes to adopt or diverting it from the evidence that seems to disconfirm it. According to Davidson (1985a: 208) this is typically what the action of deceiving oneself is about: "The action involved may be no more than an intentional directing of attention away from the evidence in favor of p; or it may involve the active search for evidence against p". To conclude, we may say that even though it seems impossible to believe something at will, *hic et nunc*, it seems nevertheless possible to induce the adhesion to certain beliefs indirectly, via the manipulation of the process of evidence gathering.

One appealing aspect of the intentionalist account is that it renders the agent responsible for his epistemic behavior. Insofar as self-deception stems from the agent's deliberate intention, just like other-deception, it makes sense to impute to him the responsibility for adopting an illusional belief. Rather than a mere victim of his illusions, the agent appears to be the author of an irrational act, someone who chooses to "bury his head in the sand" or "refuses to face reality", to use some common expressions. Further, the self-deceiver may also be held accountable for the potential implications of his act, given that illusional beliefs may have negative consequences not only for the agent but also for others. With regard to the agent himself, some studies indicate that positive illusions are often maladaptive. For example, cancer patients who are overly optimistic about their medical condition may fail to seek treatment when faced with alarming symptoms of cancer recurrence (Dunning et al. 2004). And more generally, as Taylor (1989: 237) points out, "Unrealistic optimism might lead people to ignore legitimate risks in their environment and to fail to take measures to offset those risks". With regard to the implications of self-deception for others, Clifford's famous example of the shipowner provides a very suggestive illustration. Despite suspecting that his ship was no longer safe and seaworthy, and that letting it sail would put at risk many people's lives, the shipowner persuades himself that nothing could go wrong, "put[ting] his trust in Providence, which could hardly fail to protect all these unhappy families" and "dismissing from his mind all ungenerous suspicions" (Clifford 1877: 177). The interest of this example is that it

[2] See also Davidson (1985: 208, note 5) and Mele (2001: 16) for similar examples.

emphasizes the extent to which someone's illusions may be harmful to others. And assuming that such illusions are self-induced, as intentionalists claim, we are just as responsible for irrational beliefs as we are responsible for irrational actions.

3. *The paradoxes of intentionalism*

The intentionalist hypothesis raises nonetheless several paradoxes which cast doubt on its validity. The first was brought to light by Sartre in his famous analysis of *mauvaise foi*. To be able to persuade myself that *p* is true I must initially recognize that *p* is false, otherwise I would not *deceive* myself strictly speaking, but simply commit an involuntary mistake. Sartre (1969: 49) writes: "The one to whom the lie is told and the one who lies are one and the same person, which means that I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived". But if the process of deceiving myself into believing that *p* requires me to acknowledge initially that not-*p*, it must be that I hold two contradictory beliefs simultaneously at a given point in time. But is that mentally possible? Is it possible, for example, to believe that I am losing my hair and, at the same time, believe that I am not losing my hair?

Mele (1998) points out a second difficulty, which concerns the very possibility of executing a strategy to fool oneself. He terms it the *strategy* paradox: "In general, *A* cannot successfully employ a deceptive strategy against *B* if *B* knows *A*'s intention and plan. This seems plausible as well when *A* and *B* are the same person. A potential self-deceiver's knowledge of his intention and strategy would seem typically to render them ineffective" (Mele 1998: 38). Even if the agent's strategy consists simply in disregarding evidence for the dreaded belief, the paradox remains, insofar as the effort to neglect a particular information presupposes an awareness of that very information. As Baumeister (1993: 168) observes, "one must first notice something in order to be careful not to notice it—but if one has already noticed it, then it is too late".

Furthermore, the assumption that self-deception is an intentional and deliberate process relies on the assumption that it is beneficial to adopt an illusional belief. Yet, in most cases it seems counter-productive to embrace a falsehood, insofar as it undermines the ability to intervene in the world efficiently in order to promote our goals and interests. This problem could be termed the *economic paradox of positive illusions*: on the one hand, the purpose of self-deception would be to acquire a pleasant or rewarding belief, but on the other hand the very adoption of a false belief seems to compromise the capacity to maximize well-being in the long-term. Elster (1999: 438) insists on this aspect: "To navigate the world successfully, one needs beliefs that are as good as they can be, given the available evidence. Someone who believes that *p* merely because he wants *p* to be the case is more likely to see some of his long-term goals frustrated". After all, the maximization of

our preferences depends to a great extent on an accurate assessment of (a) the foreseeable costs and benefits of each feasible option, and of (b) the probability associated to each option. If an agent is self-deceived about any of these aspects, she runs the risk of choosing an option that eventually minimizes her well-being. Thus, to come back to Carlos' example, we may observe that by getting himself to believe that he will probably succeed in the test Carlos only obtains a small benefit (reduction of anxiety, boost of self-confidence), while he risks undermining the greater benefit of passing the test, to the extent that his false sense of preparedness is likely to discourage any further efforts of preparation. A more realistic assessment, on the other hand, would allow him to acknowledge his limitations and to overcome them through his actions, thereby increasing his chances to succeed in the exam.[3]

And finally, assuming that self-deception requires the subject's intention, how can we make sense of negative cases of self-deception—what Mele (1999) proposes to call 'twisted self-deception'—that is, cases in which the subject persuades herself of something painful and undesirable. How are we to explain, for example, the jealous person's delusional belief that her partner is being unfaithful, or the paranoid's unjustified belief that she is constantly being persecuted, or the pessimist's unrealistically negative predictions of the future? Why would someone intentionally decide to adopt beliefs that have undesirable effects both in the short-term (they make us "feel bad") and in the long-term (they propel irrational actions)? This phenomenon could be termed the *economic paradox of negative illusions*.

Most proponents of intentionalism try to bypass these paradoxes by postulating the hypothesis that the mind is divided (Audi 1982, Davidson 1985b, Fingarette 1982, Gardner 1993, Pears 1984). In light of this assumption, to say that a given person deceives herself into believing that *p* actually means that *a part of* her mind, who knows (or suspects) that *p* is false, deceives another part of her mind into believeing that *p* is true. To that extent, it seems no longer paradoxical to suggest that the self-deceiver holds two contradictory beliefs, given that those beliefs pertain to different sub-systems of the mind. However, the divisionist hypothesis is an *ad hoc* postulate that arguably raises more problems than it solves[4]. Although it is widely acknowledged that there are pathological cases of 'mental dissociation' (or 'multiple personality disorders') that involve a differentiation of the subject's personality, it has never been confirmed empirically that people's minds are *essen-*

[3] The exception to this principle would perhaps be the terminally ill patient who deceives himself into believing that he will survive. But that is due precisely to the fact that the notion of 'long-term' ceases to make sense in such a context.

[4] Davidson (1985b: 353) explicitly acknowledges that the only reason he resorts to the divisionist hypothesis is that there is no other way, in his view, to account for irrational phenomena such as self-deception and akrasia: "I have urged in several papers that it is only by postulating a kind of compartmentalization of the mind that we can understand, and begin to explain, irrationality".

tially composed of different sub-systems, as Freudian accounts claim, nor that they experience temporary divisions whenever they act, think or feel irrationally. Moreover, as Wittgenstein points out, the divisionist model implies the ascription of attitudes such as desires, beliefs, representations, memories and sensations to different sub-systems of the mind when it only makes sense to ascribe these attitudes to the person as a whole (the *in*-dividual). Wittgenstein (1958: § 281) writes: "Only of a living human being and what resembles (behaves like) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious". The problem with the ascription of anthropomorphic attributes to alleged parts of the mind—which Kenny (1971: 65) calls the 'homunculus fallacy'—is that it portrays the different *homunculi* as if they were different persons within the person without being able to account for the unity of the whole.[5]

4. *The motivational account*

Davidson (1985b: 184) claims that "the underlying paradox of irrationality, from which no theory can entirely escape, is this: if we explain it too well, we turn it into a concealed form of rationality". Yet, it seems possible to overcome each of the above-mentioned paradoxes without denying the irrational nature of the phenomenon, provided that we accept the idea that self-deception is due to a particular motive, rather than to an intention, and more exactly to the influence of that motive upon the subject's cognitive faculties (Barnes 1997, Lazar 1999, Mele 1987, 2001a). According to the 'motivational' account, self-deception is typically an involuntary process that occurs without the subject's awareness, via the cognitive distortions (or biases) that emotional states are liable to exert on people's judgment. The motive in question could be the desire to believe that *p* (Mele 1987), the anxiety that not-*p* (Barnes 1997) or any other emotion related to *p* (Lazar 1999, Mele 2001a).

This hypothesis is consistent with the profusion of empirical studies carried out by cognitive and social psychologists from the 1960s onwards, which indicate that emotions may affect the way we process and interpret the available evidence in a variety of ways (for a review, see Kunda 1999; Gilovitch et al. 2002). More specifically, emotions seem to induce motivationally biased beliefs at three different levels (Mele 2001a: 26–27). Firstly, emotions affect the subject's *memory* and *attention* to evidence. An excessively jealous person, for instance, tends to focus too much on her partner's behavior and to see every unaccountable attitude as a sign of infidelity. This is why Iago finds it so easy to manipulate Othello, for he knows that "Trifles light as air are to the jealous confirmations strong as proofs of holy writ" (*Othello*, III, 3). Secondly, partly because of this, emotions also affect the process of

[5] This objection was initially raised by Sartre (1969: 53) in his critique of Freud's theory.

evidence gathering. In motivated cases of 'confirmation bias', in particular, people tend to favor the information that confirms what they desire to believe. According to Oswald and Grosjean (2004: 81), "this tendency exists … because the possibility of rejecting the [desired belief] is linked to anxiety or other negative emotions". And thirdly, emotions may also bias the very *interpretation* of the available information. A typical example of this is the person 'in love' who mistakes a sign of friendship for the expression of a mutual feeling.

The motivational approach presents three main advantages in comparison to the intentionalist approach. In the first place, as we have seen, it fits well with the empirical studies conducted by psychologists in the past decades. Although these studies do not falsify the intentionalist hypothesis strictly speaking, they confirm that the motivational hypothesis is valid for a vast number of cases. In the second place, it provides a *generalized theory of cognitive irrationality*, insofar as it refers to one and the same psychological phenomenon—motivated judgment—to account not only for self-deception, but also for wishful thinking, rationalization and a host of other motivated biases (confirmation bias, egocentric bias, optimism bias, overconfidence effect, self-serving bias, etc).

And in the third place, perhaps more importantly, the motivational model seems to overcome each of the above-mentioned paradoxes—which no longer appear to be the so-called "paradoxes of self-deception", but more exactly the *paradoxes of the intentionalist account of self-deception*. In fact, once we assume that self-deception is an involuntary and unconscious process, the two first paradoxes cease to pose a threat: On the one hand, the assumption that two contradictory beliefs must coexist in the subject's mind is no longer required (doxastic paradox), given that the discrepancy lies solely between the subject's belief and what the evidence clearly suggests; and, on the other hand, the conundrum of a subject who intentionally deceives himself without being aware of his own intention and plan of deception (strategy paradox) also seems to dissipate, given that the subject is now seen as the victim and not the author of the illusion. Similarly, the two economic paradoxes also disappear if we accept the idea that self-deception stems from the influence of emotions on cognitive processes. Unlike practical reasoning and intentional decisions, motivational biases are not supposed to ensure the maximization of people's well-being, neither in the long-term (paradox of positive cases) nor in the short-term (paradox of negative cases). And finally, the motivational hypothesis does not require the postulate that the mind is divided, thus avoiding the homunculus fallacy, given that the conflict within the subject's mind is polarized between an emotion (or a desire) and a belief, and not between two contradictory beliefs. This is not to deny that mental partitioning may occur in extreme cases of irrationality, but simply that the divisionist hypothesis is *not required* to account for ordinary cases of self-deception.

If this analysis is correct, the subject who deceives herself (or perhaps more accurately, the subject who is *self-deceived*) is simply the victim of a phenomenon of judgment distortion that is both involuntary and unconscious (Kunda 1990, Pohl 2004, Mercier & Sperber 2009). However, it is worth noting that the motivational account is not incompatible with the notion of *epistemic responsibility*, that is, the notion that people are responsible for the way they think, and not just for the way they act (Audi 2008, Clifford 1877, Engel 2001). Granted, beliefs are typically involuntary states (Bennett 1990, Montmarquet 2008, Williams 1973), and so is the phenomenon of motivated irrationality. However, we have the ability to exert a certain degree of control over the process of belief formation. As Audi (2008: 403) points out, we must distinguish between the question of the *voluntariness of belief*, on the one hand, and the question of the *voluntariness of the grounding of belief*, on the other. Although one cannot avoid self-deception directly, by an act of will, given that the process occurs without people's awareness, it seems possible nonetheless to make sure that our beliefs are formed in conformity with certain epistemic requirements (e.g., consistency, justification, consideration of all available evidence). The greater the epistemic control over the process of belief formation, the lesser the impact of emotions on the cognitive processes. In this sense, it seems reasonable to conclude that we are *partly* and *indirectly* responsible for our illusions.

5. *Self-deception, akrasia and precommitment*

The question of epistemic responsibility appears all the more important if we take into account the impact that irrational beliefs seem to have on the practical sphere. More specifically, the phenomenon of motivated irrationality is also susceptible to affect the *evaluative judgments* through which we assess the value of feasible options, and subsequently the optimality of the decisions that are based upon those judgments.

According to some decision theorists (Ainslie 2001, Elster 2007; see Loewenstein et al. 2003 for a review) this is precisely what happens when people act against their better judgment (*akrasia*). The weak-willed agent is generally described as a person who judges that the option A (e.g. improve her health) is clearly preferable, all things considered, to the option B (e.g. smoke a cigarette); decides to do A rather than B after a pondered deliberation, and ends up doing B nevertheless. The most paradoxical aspect about this phenomenon is that the weak-willed agent chooses the worst available option despite knowing that it minimizes her interest. One plausible explanation, however, is that the akratic agent falls prey to a cognitive bias—known as the "discounting bias" in the language of psychologists—which may be characterized as a sort of "myopia" regarding future preferences (Strotz 1956). Loosely speaking, the idea is simply that we tend to overrate the instrumental

value of immediate options and, conversely, to underrate the instrumental value of future options. As Ainslie (2001: 38) observes, "[people] tend to prefer smaller, earlier rewards to larger, later ones temporarily, during the time that they're imminent". In addition, Ainslie's research shows that the curve describing the devaluation of future goods proportionally to their delay is *hyperbolic*, and not just exponential, which means that "the smaller reward is temporarily preferred for a period before it is available" (Ainslie 2001: 32). In consequence, when the motivation to defer gratification is insufficient, we may end up giving in to temptation and choosing a small immediate reward instead of a future greater reward.

To that extent, the problem is not that the agent acts against his own judgment (Davidson 1980, 1985b), but more exactly that he acts on the basis of a temporarily biased judgment, presumably due to the influence of a strong emotion or desire. In this sense, as Elster (1999: 429) points out, akrasia typically involves a *diachronic inconsistency* (rather than a synchronic inconsistency) between the agent's attitudes. For instance, an agent who has decided to loose weight may enter a restaurant at time 1 with the certainty that it is preferable to sacrifice the small reward of having a dessert in order to ensure the greater reward of loosing weight with health benefits. Yet, Elster explains, the imminent availability of the reward may affect his judgment for a moment, and during that moment the reward of eating a dessert might appear superior to the remote reward of being healthier: "As the meal progresses, a preference reversal occurs at time t*, and when the waiter asks him, at time 2, whether he wants to order dessert he answers in the affirmative … He is not, however, acting against his better judgment *at the time of ordering dessert*" (Elster 1999: 430). After leaving the restaurant—say, at time 3—it is likely that the agent returns to his initial preference, acknowledging with regret that he made a mistake. But that is presumably due to the fact that desires fade once they are satisfied, and so do their effects on people's judgment. Once the craving for the dessert dissipates, the agent is able to see clearly that it was indeed in his best interest to abstain from the immediate pleasure.

At the same time, this understanding of the mechanisms underlying akratic behaviour seems to have considerable implications on the normative level. In particular, the awareness that the propensity to succumb to temptation is often due to the influence of desires on practical judgement—what one might call a "weakness of the judgment" rather than a "weakness of will"—facilitates the elaboration of indirect methods of self-control. For example, the person who wants to loose weight may adopt the strategy of buying groceries shortly after a meal, when cravings are weaker and therefore less likely to trigger a preference reversal. Another efficient strategy consists in limiting future options deliberately in an attempt to eliminate the very possibility of succumbing to temptation (Elster 2007: 237). An obvious method to avoid

overeating sweets at home, for example, is to avoid having sweets at home. Likewise, the indebted consumer who systematically succumbs to the temptation of purchasing on credit might want to cut up his credit cards with a pair of scissors. A similar strategy is proposed in some countries to pathological gamblers who wish to overcome their addiction: by signing a voluntary self-exclusion declaration, the gambler is irreversibly banned from casinos and can no longer succumb to the urge of playing again. This sort of device corresponds to what economists generically term *precommitment strategy*, which consists in eliminating or imposing restrictions on future options. The *locus classicus* of precommitment is the Homeric episode in which Ulysses instructs his crewmen to tie him to the mast to be able to hear the Sirens' alluring songs without incurring the risk of running his ship onto the treacherous rocks. Precommitment thus involve the acknowledgement of one's propensity to give in to temptation, along with the notion that the "present self" should constraint the choices of the "future self" in order to ensure the maximization of well-being in the long-term.

But self-control by precommitment does not always entail the elimination of future options. In certain cases, it may be more useful to impose a sanction on the tempting option. Elster (2007: 238) evokes a convincing example: "If I begin saving for Christmas but find myself taking money out of my savings account instead of keeping it there … I may put my savings into a high-interest account that carries a penalty for early withdrawal, thus combining premium and penalty". In a study about the efficiency of precommitment against procrastination, Arieli and Wertenbroch (2002: 221) demonstrated that "people are willing to self-impose deadlines to overcome procrastination, even when these deadlines are costly". More significantly, their study showed that precommitment to deadlines is successful both in reducing procrastination and in helping students achieve better grades. For example, students who chose to be penalized at the rate of one percent of the grade for each day late had in average better marks than their peers. As Arieli (2009: 116) later pointed out, these results are interesting in that they suggest that "although almost everyone has problems of procrastination, those who recognize and admit their weakness are in a better position to utilize available tools for precommitment and by doing so, help themselves overcome it".

Moreover, the motivational model also predicts the efficiency of self-control strategies based on emotional regulation. Assuming that preference reversal is generally caused by a strong affect, managing emotions and moods appears to be one of the best methods to forestall impulsive behaviour. Although emotions, like beliefs, are in principle involuntary states, it seems possible to achieve this indirectly, either by focusing on the pertinent stimulus or by manipulating the external conditions (Mikolajczak et al. 2009: 168). The employee who doubts he will have enough courage to ask his boss for a raise may try to recall

vividly the injustices that he has suffered at work (Skinner 1953: 236). The mother who dreads the consequences of feeling overly angry at her child may attenuate that anger by focusing on the good moments they spent together (Mele 2001b: 106). And the person who is about to commit suicide may think of her beloved ones, or have the lucidity of thinking that it is probably a negative emotion that triggers an unrealistically pessimistic vision of the future. In each of these examples, the key aspect is that the agent's strategy of self-control relies on the acknowledgement that her future judgment may be prone to motivational biases.

Finally, it is important to stress that these and other self-control strategies may also prove useful within the epistemic realm. It was suggested in the previous section that we are partly responsible for the irrationality of our beliefs, to the extent that we may exert a certain degree of control over the process of belief formation. This claim has two significant implications. On the one hand, as we have seen, it becomes possible to develop what some authors call *doxastic self-control* (Audi 2008, Mele 2001b) in an effort to counteract the effect of motivational biases. In this sense, as Mele (2001b: 98) observes, self-control strategies may profitably be adopted to ensure the rationality of our beliefs, and not just of our actions: "Self-control, then, extends beyond action to belief". But conversely, it becomes apparent that doxastic self-control is paramount, in turn, to ensure the rationality of our actions. The ability to prevent akratic behaviour, in particular, may presumably be constrained by our efforts to counteract motivational biases. This may be achieved through the adoption of reliable methods of inquiry, or through the development of a certain number of "intellectual virtues" (Zagzebski 1996, Montmarquet 2008). But precommitment strategies may also prove useful in the epistemic sphere. For example, a judge who recognizes that his racial prejudices are susceptible to influence his judgment may scrupulously resort to the precommitment strategy of forcing herself to go through every piece of evidence once again before reaching a sentence. Likewise, the financial investor who is aware that optimism biases often lead people to overestimate the likelihood of positive events may self-impose a certain number of restrictions on all future investments (e.g., never invest more than what I can afford to lose, never put all my eggs in the same basket, always allocate my money equally to each of N funds). Whatever the strategy, the first step to ensure the rationality of our attitudes, whether on the practical or epistemic level, is the acknowledgment of the fallibility of human judgment under the influence of emotions.

6. *Conclusion*

This paper showed that an adequate understanding of self-deception holds the key to developing efficient strategies of self-control both in the practical and in the cognitive sphere. While intentionalists suggest

that self-deception is a paradoxical phenomenon, we have seen that the alleged paradoxes characterize the intentionalist model itself. In contrast, motivational models provide a non-paradoxical account of irrationality without appealing to the postulate that the mind is divided[6]. This model proves consistent with the empirical studies carried out by social psychologists in the past decades, which converge in demonstrating that most people fall prey systematically to a variety of cognitive and motivational biases. Moreover, insofar as such biases also have an impact on practical judgments, and subsequently on our actions, it seems reasonable to speak of a unified account of irrationality, according to which self-deception and akrasia are rooted in the phenomenon of motivated irrationality, that is, in the influence that emotions and desires exert over people's judgment. But the fact that motivational biases are both unconscious and involuntary does not seem to imply that people are merely the victims of their own illusions. We have the possibility, and perhaps the moral obligation, to adopt indirect strategies of doxastic self-control. If correctly implemented, these strategies have the ability to promote the rationality of our beliefs and our actions.

References

Ainslie, G. 2001. *Break-down of the Will*. Cambridge: Cambridge University Press.

Arieli, D. & Wertenbroch, K. 2002. "Procrastination, deadlines, and performance: Self-control by precommitment". *Psychological Science* 13 (3): 219–224.

American Psychiatric Association 2000. *Diagnostic and Statistical Manual of Mental Disorders*. 4th Edition. Washington: American Psychiatric Association.

Arieli, D. 2009. *Predictably Irrational*. London: Harper Collins.

Audi, R. 1982. "Self-Deception, Action and Will". *Erkenntnis* 18: 133–158.

Audi, R. 1988. "Self-deception, rationalization, and reasons for acting". In A. Rorty & McLaughlin. *Perspectives on Self-Deception*. Berkeley: University of California Press.

Audi, R. 2008. "The ethics of belief: Doxastic self-control and intellectual virtue". *Synthese* 161: 403–418.

Barnes, A. 1997. *Seeing Through Self-Deception*. Cambridge: Cambridge University Press.

Baron, J. 1988. *Thinking and Deciding*. Cambridge: Cambridge University Press.

Baumeister, R. 1993. "Lying to Yourself: The enigma of self-deception". In Lewis & C. Saarni (eds.). *Lying and Deception in Everyday Life*. New York, London: Guilford Press.

Bennett, J. 1990. "Why is believing involuntary?". *Analysis* 50: 87–107.

[6] This is not to deny that there are cases of mental dissociation, but they are generally pathological and involve a "dissociative identity disorder" that has little to do with ordinary cases of self-deception (American Psychiatric Association 2000: 529).

Bermudez, J.L. 1997. "Defending Intentionalist Accounts of Self-Deception". *Behavioral and Brain Sciences* 20: 107–108.

Clifford, W. 1994. "The Ethics of Belief". In Klemke, Kline and Holinger (eds.). *Philosophy: Contemporary Perspectives on Perennial Issues*. New York: St. Martin's Press.

Davidson, D. 1980. "How is Weakness of Will Possible?" In *Essays on Actions and Events*. Oxford: Oxford University Press.

Davidson, D. 1982. "Paradoxes of Irrationality". Reed. in *Problems of Rationality*. Oxford, New York: Clarendon Press, 2004.

Davidson, D. 1985a. "Deception and Division". Reed. in *Problems of Rationality*. Oxford, New York: Clarendon Press, 2004.

Davidson, D. 1985b. "Incoherence and Irrationality". *Dialectica* 39: 345–354.

Dunning, D., Heath, C. & Suls, J. M. 2004. "Flawed self-assessment: Implications for health, education, and the workplace". *Psychological Science in the Public Interest* 5: 69–106.

Elster, J. 1979. *Ulysses and the Sirens*. Cambridge: Cambridge University Press.

Elster, J. 1999. "Davidson on weakness of will and self-deception". In L. E. Hahn (ed.). *The Philosophy of Donald Davidson*. Open Court: La Salle: 425–41.

Elster, J. 2007. *Explaining Social Behavior*. Cambridge: Cambridge University Press.

Engel, P. 2001. "Sommes-nous responsables de nos croyances?" In Y. Michaux (ed.). *Qu'est-ce que la culture?* Paris: Odile Jacob.

Fingarette, H. 1982. "Self-deception and the splitting of the ego". In R. Wollheim and J. Hopkins (eds.). *Philosophical Essays on Freud*. Cambridge: Cambridge University Press.

Gardner, S. 1993. *Irrationality and the Philosophy of Psychoanalysis*. Cambridge: Cambridge University Press.

Gilovich, T. 1991. *How We Know What Isn't So*. New York: The Free Press.

Gilovitch, T. Griffin, D. & Kahneman, D. 2002. *Heuristics and Biases*. Cambridge: Cambridge University Press.

Heil, J. 1984. "Doxastic Incontinence". *Mind* 93: 56–70.

Johnston, M. 1988. "Self-Deception and the Nature of Mind". In A. Rorty & McLaughlin. *Perspectives on Self-Deception*. Berkeley: University of California Press.

Kenny, A. 1984. *The Legacy of Wittgenstein*. Oxford: Basil Blackwell.

Kunda, Z. 1990. "The Case for Motivated Reasoning". *Psychological Bulletin* 108 (3): 480–498.

Kunda, Z. 1999. *Social Cognition. Making sense of people*. Cambridge: MIT Press.

Lazar, A. 1999. "Deceiving Oneself or Self-Deceived? On the Formation of Beliefs Under the Influence". *Mind* 108: 265–290.

Loewenstein, G., Read, D. and Baumeister, R. (eds.). 2003. *Time and Decision*. New York: Russell Sage Foundation.

Martin, M. 2009. "Happily Self-Deceived". *Social Theory and Practice* 35 (1): 29–44.

McKay, R. T. & Dennett, D. 2009. "The Evolution of Misbelief". *Behavioral and Brain Sciences* 32: 493–561.

McLaughlin, B. 1988. "Exploring the possibility of self-deception". In A. Rorty & McLaughlin. *Perspectives on Self-Deception*. Berkeley: University of California Press.

Mele, A. 1987. *Irrationality*. New York, Oxford: Oxford University Press.

Mele, A. 1998. "Two Paradoxes of Self-Deception". In J-P. Dupuy (ed.). *Self-Deception and Paradoxes of Rationality*. Stanford: CSLI Publications.

Mele, A. 1999. "Twisted Self-Deception". *Philosophical Psychology* 12: 117–137.

Mele, A. 2001a. *Self-Deception Unmasked*. Princeton, Oxford: Princeton University Press.

Mele, A. 2001b. *Autonomous Agents*. Oxford, New York: Oxford University Press.

Mercier, H. & Sperber, D. 2009. "Intuitive and reflective beliefs". In J. Evans and K. Frankish (eds.). *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Mikolajczak, M., Quoidbach, J., Kotsou, I. & Nélis, D. 2009. *Les compétences émotionnelles*. Paris: Dunod.

Montmarquet, J. 2008. "The voluntariness of virtue and belief". *Philosophy* 83 (3): 373–390.

Naylor, M. 1985. "Voluntary Belief". *Philosophy and Phenomenological Research* 45: 427–436.

Oswald, M. & Grosjean, S. 2004. "Confirmation bias". In R. F. Pohl (ed.). *Cognitive Illusions*. Hove, New York: Psychology Press.

Pascal, B. 1958. *Pensées*. New York: E.P. Dutton & Co.

Pears, D. 1984. *Motivated Irrationality*. Oxford: Clarendon Press.

Pohl, R. F. (ed.). 2004. *Cognitive Illusions*. Hove and New York: Psychology Press.

Rorty, A. 1983. "Akratic Believers". *American Philosophical Quarterly* 20: 175–183.

Sartre, J-P. 1969. *Being and Nothingness*. London: Routledge Classics Series.

Scott-Kakures, D. 1996. "Self-deception and internal irrationality". *Philosophy and Phenomenological Research* 56: 31–54.

Strotz, R. H. 1956. "Myopia and inconsistency in dynamic utility maximization". *Review of Economic Studies* 23: 165–180.

Taylor, S. E. 1989. *Positive Illusions: Creative Self-Deception and the Healthy Mind*. New York: Basic Books.

Twerski, A. J. 1997. *Addictive Thinking*. Center City, MN: Hazelden.

Williams, B. 1973. "Deciding to Believe". In J. Elster (ed.). *Problems of the Self*. Cambridge: Cambridge University Press.

Wittgenstein, L. 1958. *Philosophical Investigations*. Oxford: Blackwell.

Zagzebski, L. 1996. *Virtues of the Mind*. Cambridge: Cambridge University Press.