teorema
Vol. XXVI/3, 2007, pp. 45-58
ISSN: 0210-1602

Cognitive Bias, Intentionality and Self-Deception

Anna Nicholson

RESUMEN

El presente artículo investiga el papel de la intención en el autoengaño. Como es sabido, Alfred Mele ha defendido que en ciertos casos de autoengaño, contrariamente a lo que mantiene la concepción heredada, no es necesario que se atribuya al agente la intención de engañarse a sí mismo. En contra de la posición de Mele, este ensayo destaca, en primer lugar, algunas debilidades teóricas de su postura y, en concreto, critica la función que asigna a los sesgos cognitivos en el autoengaño. En segundo lugar, perfila una concepción alternativa del fenómeno que, manteniendo la estructura de la creencia de la posición de Mele, recaracteriza el mecanismo del sesgo cognitivo y el papel de la intencionalidad con el fin de proporcionar una concepción más intuitiva del autoengaño.

ABSTRACT

This article investigates the role of intention in self-deception. Alfred Mele has famously argued that contrary to the received view, in typical instances of self-deception the agent need not be attributed the intention to self-deceive. I argue against Mele by first pointing out the theoretical weaknesses of his account, particularly criticizing the function he assigns to cognitive biases in self-deception. I then sketch an alternative view of the phenomenon which, while retaining the belief-structure of Mele's account, recharacterizes the cognitive biasing mechanism and the role of intentionality to give a more intuitive account of self-deception.

It is commonly accepted that humans have the capability to deceive themselves, yet upon closer inspection, the paradoxical nature of self-deception makes it seem difficult to understand how the phenomenon actually occurs. For instance, questions arise as to how self-deceptive beliefs come about and are maintained, and how self-deceivers fail to assimilate evidence and formulate beliefs in a way that seems obvious to the impartial observer. Alfred Mele argues that most theorists "have made self-deception more theoretically perplexing than it actually is by imposing upon the phenomena a problematic conception," [Mele (1997), p. 100] and that there is no compelling reason to implicate the agent's intention to self-deceive in typical

cases of self-deception. I will argue that there are several functional weaknesses in his description of the phenomenon, and will propose an alternative picture which attempts to address these issues while retaining the global form and doxastic structure of Mele's account.

I. Paradigm Cases of Self-Deception

Paradigmatic cases of self-deception are useful in examining the theoretical constructs that have been proposed to help explain them. The first case involves an agent who is self-deceived in maintaining a belief, and the second concerns an agent who is self-deceived in acquiring a belief.

The first is a classic case of self-deception offered by Mele in various incarnations [Mele (1987a), (1997)]. It fits neatly into the stereotypical notion of the phenomenon — the person who refuses to accept that his or her spouse is unfaithful:

> Sam has believed for many years that his wife, Sally, would never have an affair. In the past, his evidence for this belief was quite good. Sally obviously adored him; she never displayed a sexual interest in another man; [. . .] she condemned extramarital sexual activity; she was secure, and happy with her family life; and so on. However, things recently began to change significantly. Sally is now arriving home late from work on the average of two nights a week; she frequently finds excuses to leave the house alone after dinner; and Sam has been informed by a close friend that Sally has been seen in the company of a certain Mr. Jones at a theater and a local lounge. Nevertheless, Sam continues to believe that Sally would never have an affair. Unfortunately, he is wrong. Her relationship with Jones is by no means platonic [Mele (1987a), pp. 131-132].

In this case, Sam seems self-deceived in maintaining his belief that Sally is a faithful wife. Similar scenarios include the mother who does not believe that her son is a drug addict despite his behavioral indications and his room full of paraphernalia, or the wife who refuses to believe that her terminally ill husband will die, despite a preponderance of evidence to the contrary.

The second paradigm case [adapted from Mele (1997)] involves a young boy, Tom, whose father has recently died of alcoholism. He finds comfort in thinking back to some enjoyable times spent hiking and fishing with his father, gazing at photographs of the two of them posing happily together, reminiscing about playing catch in the yard, and other such pleasant and comforting memories. In reality, however, these happy moments were few and far between. He finds it upsetting to focus on the many times his father came home late at night in a drunken stupor, disappeared for days on end, missed important family events, shouted at Tom and his mother, and many other similar instances of neglect. However, in time Tom comes to be-

lieve that he had a kind, loving and attentive father. The intuitive folk-concept of self-deception would classify Tom as self-deceived in holding this belief.

II. The Paradox of Self-Deception

The first step in understanding self-deception is to address its inherent paradox, which arises when it is conceptually pinned to the construct of other-deception. In cases of other-deception, person A believes ¬p and *intentionally* causes person B to believe that-p. Intentionality on the part of the deceiver and the deceived's ignorance of the deception are both necessary conditions. When the model of other-deception is strictly applied to the self-deceived, the subject must intentionally deceive himself into holding a belief while concurrently believing its negation, resulting in the following conditions for self-deception:

A) The subject holds two contradictory beliefs simultaneously: the true belief that p and the false belief ¬p.

B) The subject intentionally causes himself to believe ¬p despite the preponderance of evidence for p.

These conditions give rise to two separate paradoxes. The consequence of condition (A) is known as the static or doxastic paradox[1]. As both 'deceiver' and 'deceived', the subject must believe that-p while at the same time believing ¬p, which is impossible. Condition (B) results in the dynamic /strategy paradox, stemming from the agent's intentionally causing himself to believe a proposition while also believing its negation. This is paradoxical because the agent must actively persuade himself to believe a proposition which he knows is false, a maneuver that is successful only if he is simultaneously aware of it (as the deceiver) and unaware of it (as the deceived). This is the critical stumbling block to understanding how self-deception is possible for a rational agent.

III. Accounting for the Paradox. Intentionalists vs Non-Intentionalists

Philosophical accounts of self-deception can be categorized according to how closely they mirror the model of other-deception. Those faithful to the model of other-deception typically postulate some type of mental partitioning in order to deal with the doxastic paradox. This partitioning facilitates the subject's ability to effectively deceive himself without being aware of his own strategic intention to do so. Such strategies range from the extreme divisionism of Rorty (1972, 1988) to the more moderate partitioning strategies of

Pears (1985) and Hamlyn (1971), but are all able to retain the subject's intention — with respect to one of the partitions — to self-deceive, and are thus classified as intentionalist approaches.

Other philosophers avoid mental partitioning by rejecting some element of the other-deception model. Unitary approaches address the doxastic paradox by arguing that the self-deceived subject need not believe both p and ¬p. However, they differ with respect to which of those beliefs the subject actually holds. Audi (1982) and Whisner (1998) contend that the subject holds the true belief, but not the false one. Others argue that the subject holds the false belief, but not the true one [e.g. Canfield and Gustavson (1962), Johnston (1988), Lazar (1999), and Mele (1987b, 1997, 2003)]. These types of approaches are usually anti-intentionalist as well, rejecting the notion that typical instances of self-deception are conditional upon the agent's intention to self-deceive.

IV. MELE'S APPROACH

Mele's account of self-deception is characteristic of those which consider the acquisition or maintenance of a self-deceptive false-belief to be the result of some sort of motivationally (but not intentionally) biased cognition, usually triggered by emotion or desire. Mele concedes that intentional self-deception is possible[2] in contrived cases, but contends that "garden-variety" instances of self-deception can be explained without requiring that the agent have the intention to self-deceive. Because the subject need only hold the false-belief, and not its contradiction, Mele also avoids any mental partitioning. As motivated behavior is not coextensive with intended behavior, he argues that motivational states can trigger biased intention without the agent's intention to do so. Typical strategies of the self-deceptive agent involve either the control of new input, or the internal biasing of information already acquired. The following conditions are offered as jointly sufficient (but not necessary) for entering self-deception [Mele (1997), p. 95]:

> The belief that *p* which S acquires is false[3].
>
> S treats data relevant, or at least seemingly relevant, to the truth value of *p* in a motivationally biased way.
>
> This biased treatment is a non-deviant cause of S's acquiring the belief that *p*.
>
> The body of data possessed by S at the time provides greater warrant for ¬*p* than for *p*.

Or more simply put, that "people enter self-deception in acquiring a belief that-p if and only if p is false and they acquire the belief in a suitably biased way" [Mele (2003), p. 163]. Mele (1997), p. 94, outline several 'suitably' biased mo-

tivated mechanisms by which the subject's desire that p can lead to the belief that p:

Negative Misinterpretation	subject's desire that-p leads to ignoring evidence against p that would be accepted in absence of wanting p to be true
Positive Misinterpretation	subject's desire that-p leads to interpreting as supporting p evidence that would count against p in that desire's absence
Selective Focusing/Attending	subject's desire that-p leads to a lack of attentional focus on evidence against p and attentional focus on evidence supporting p
Selective Evidence Gathering	subject's desire that-p leads to ignoring evidence against p and over-focusing on evidence for p

Mele's (1997) assessment of the case of Sam and Sally provides an illustration of each of these motivated biasing mechanisms. Sam could negatively misinterpret the available evidence by asking her for a non-incriminating explanation for her actions, even going so far as to provide a reasonable excuse for her approval. He could positively misinterpret the data by deducing that if she were actually having an affair with Jones, then she would do such an effective job of concealing it that she would never be seen in public with him, thus their public meetings count against their involvement. Selective focusing and evidence gathering strategies allow Sam to ostensibly search for damning evidence, missing the obvious clues in favor of less readily accessible evidence confirming his desired belief. Despite the evidence against Sam's belief that Sally is not having an affair, his ability to maintain this belief is wholly independent of an intention to deceive himself.

Mele also argues that emotions can trigger biasing mechanisms that are usually considered to function independently of motivation, such as vividness of information and confirmation bias. As the causes of belief-biasing desires, emotions can influence both the occurrence and the salience of available hypotheses. The desire that-p can influence the agent to test the hypothesis that p is true, rather than the hypothesis that p is false. Returning to Sam and Sally, Sam's emotions — particularly, his love for his wife coupled with his fear of losing her — trigger his desire that she is not having an affair, thereby increasing the salience of any evidence that she is faithful. He thus favours testing the hypothesis that she is faithful, rather than that she is not [Mele (2003)].

With respect to the strategy paradox, Mele differentiates the types of cognition that can contribute to a motivationally biased self-deceptive belief as intentional activities, unintentional activities, and intentional activities en-

gaged in with the intention to deceive oneself. Tom's case is useful in clarifying this important distinction, as the biasing mechanisms employed by Sam would be classified as generally unintentional by Mele. Tom intentionally focuses his attention on happy memories and intentionally avoids lingering on memories of his father's regular neglect or cruelty. He intentionally surrounds himself with the relatively scarce evidence (e.g. photographs, gifts) which suggests that his father was kind and loving. As a result of such intentional cognitive activities, Tom comes to acquire the false and self-deceptive belief that his father was a loving and attentive man. However, Mele argues strenuously that in such garden-variety cases of self-deception, these intentional activities need not reflect the subject's underlying intention to deceive himself [1997].

V. Garden-Variety Self-Deception or Garden-Variety Wishful Thinking?

Different tactics for disentangling wishful thinking from self-deception are at the core of the struggle between intentionalists and anti-intentionalists. Both are types of cognitively biased belief formation[4]. The subject's evaluation of available relevant evidence is skewed toward the favored belief, which is synchronically and causally sustained by a desire [McLaughlin 1988]. Neither phenomenon requires that the false belief held by the agent be epistemically warranted by available evidence. Predictably, intentionalists argue that it is the agent's intention to self-deceive which distinguishes the two. Davidson (1986) discriminates between wishful thinking and self-deception by arguing that the former consists of a subject's desire that-p, which triggers a corresponding belief that-p, whereas the latter entails an active and intentional effort to bring about a modification of beliefs currently held by the subject. Anti-intentionalists have a variety of responses supporting their rejections of the self-deceived agent's intentionality, arguing that self-deceivers are subject to a greater balance of evidence against their favored beliefs than are wishful thinkers, that they in fact recognize the counter-evidence [Johnston (1988)], or that the self-deceiver is actually aware of his logical inclination to believe ¬p [McLaughlin (1988)].

I concur with the intentionalists that it is in fact intentionality that separates the wishful thinker from the self-deceiver, and thus Mele's construal of Tom and Sam as self-deceivers does not convincingly distinguish them from wishful thinkers. They both hold false-beliefs that are not epistemically warranted by available evidence; those beliefs are both triggered and sustained by desire. They must be aware of this evidence against their desired beliefs, because it must have been processed on some level in order to have been rejected during the evidence-filtering process. The preceding conditions hold

for both wishful thinking and self-deception, as described by Mele. There seems to be no deciding factor that identifies Tom and Sam as self-deceivers.

Furthermore, it is not entirely clear how an agent's intentional cognitive activities, which contribute directly to his entering self-deception, can really be considered not guided by an intention to self-deceive in any sense. Mele defines intentions as executive plans, with 'executive' referring to the idea that intending to do a thing means being settled upon doing it, whereas desiring or being motivated to do a thing does not imply such settledness [Mele (1997)]. In light of this, the subject who is presumably aware of his desire to believe that-p, and intentionally carries out cognitive activities (selective evidence gathering, etc) in pursuit of that goal, seems to be acting in a goal-directed and settled manner. Lazar (1997) also contends that Mele's line of argument falls too closely in line with the intentionalist accounts against which he situates his own anti-intentionalist one, in that he actually implicates the agent's desire *to form* the belief that-p, rather than the desire that-p, in triggering the motivated biasing mechanisms. Mele does not effectively defend his argument that a subject's self-deceptive belief that-p, originating in a desire that-p and executed by a host of intentional cognitive activities, is functionally different to cognitive bias triggered by intention to believe that-p. Other than his subject's not needing to believe both that-p and ¬p simultaneously, his account seems to echo the intentionalist approach with respect to the strategy paradox.

VI. The Sorting Problem

Another problematic element of Mele's approach concerns the way in which the subject comes to believe that p despite encountering a significant body of evidence in support of ¬p. It arises upon close inspection of the function and objective of the motivated cognitive biasing mechanisms, both intentional and non-intentional. As outlined in section IV, the way in which Mele's subject enters into self-deception can be summarized in the following manner:

> The subject's desire that-p (a false-belief), which is not epistemically warranted, triggers *either* an intentionally motivated cognitive bias (e.g. selective attending/evidence gathering) *or* a non-intentionally motivated cognitive bias (e.g. confirmation bias, positive/negative interpretation, selective attending/evidence gathering) which results in the subject's acquisition or maintenance of the belief that-p [Mele (1997), pp. 94-95].

Whether intentional or non-intentional, the mechanism is biased toward encountering and/or interpreting evidence in support of p. The crucial underlying problem is specifically how this biasing mechanism functions with respect to selective attending and selective evidence gathering. The mechanism is charged with somehow filtering each item of available evidence, and then

classifying each interpretation and its potential implications as supporting either p or ¬p. In other words, evidence must be processed and (at least preliminary) inferential conclusions drawn in order for it to be assessed, then accepted or rejected for attentional focus. This 'sorting problem', as I call it, refers to the notion that filtering evidence requires that even rejected stimuli be processed on some level, and to the potential effect of that processing on the subject's cognition.

The task of evidence sorting would require a level of cognition for which a non-intentional mechanism of cognitive bias would be ill-equipped. That is, the categorization of available stimuli into those to be attended to and those to be rejected or ignored seems too complex to occur without any sort of conscious intention to do so, or awareness on some level, in order to facilitate the filtering process. However, though the intentional cognitive biases of selective attending and evidence gathering (e.g. Tom's intentional focus on positive memories and photographs of his father) should be able to handle the complexity of the filtering process, they generate their own set of difficulties with respect to the sorting problem.

There is relevant experimental evidence and strong theoretical literature from cognitive psychology regarding intentional inattention and control failure. With respect to the former, Marsh et al [forthcoming] found that when experimental participants were asked to intentionally engage with stimuli presented visually (e.g. categorization tasks) while deliberately ignoring concurrent auditory stimuli, intention-relevant but ostensibly ignored information was effectively primed for recall in subsequent memory tasks. Regarding control failure, mental control refers to an agent's ability to intentionally regulate thoughts and behavior, and will be discussed further in the next section. Central to the concept of mental control are mental schema, or general constructs that help us organize knowledge about the world, and aid in the efficient interpretation of stimulus input[5]. Ironic processes of mental control refer to the unwanted effects of control failure, which tend to occur in behavior or activities which have conscious processes devoted to them [Wegner (1994)]. Ironic processes are suggested to account for distraction, obsessive or repetitive behavior (in which a stimulus activates an extremely strong automatic schema), or 'choking' — the inability to perform successfully in a performance or athletic event. It can result in the agent's inability to 'block' a certain thought. For example, a person attempting to quit smoking or lose weight will often think constantly about smoking or eating. In all of these cases, the agent's attempt to avoid or prevent a particular thought results in its being primed and highly accessible for recurrence. Similarly, the phenomenon of directed forgetting occurs when an intentional effort to forget something, for example an unpleasant memory, results in its being primed for recurrence.

In light of this, the filtering process required for an intentional cognitive bias toward encountering or interpreting evidence in support of the belief

that-p could result in an entirely opposite effect: the rejected (¬p-supporting) evidence being primed for recurrence. Returning to Mele's assessment of Tom, his determination to encounter and interpret information concerning his father as supporting his desired belief was manifested in a conscious intention to bias his cognition to that end. However, the evidence from cognitive psychology would suggest that his strategy is in fact counterproductive. The filtering necessary to sort the evidential input will require that the rejected stimuli are processed to some degree, thus the mental schemata associated with the unwanted stimuli are strengthened, primed and highly accessible for future recurrence. The same unwanted consequence would also occur as a result of Tom's intentionally directed attempts to forget bad memories. Despite Mele's [Mele (1997), p. 99] argument that Sam's consciously-held intention to self-deceive would "undermine the project," it seems that his consciously-held intention to bias his cognition, not to self-deceive, would be no less ineffectual.

Eliminating both the element of intentionality and the requirement that the subject hold the simultaneous beliefs that-p and ¬p strays too far from the model of other-deception, upon which both the folk- and philosophical concepts of self-deception are based. There is something intuitively dissatisfying about removing both of those conditions from the formula, and continuing to call its product self-deception. I would argue that it is theoretically possible to introduce the agent's intention to self-deceive at the initial stages of the process of self-deception, without necessitating any form of mental partitioning, thus maintaining an anti-divisionist stance against the doxastic paradox. Shifting the functional objective of the subject's cognitive bias in conjunction with introducing an intention to self-deceive could also serve to eliminate the 'sorting problem' of filtering out positive and negative evidence. The next section will address how these changes might be structured.

VII. An Alternative Picture

The alternative picture I propose involves introducing the agent's intention to self-deceive in garden-variety instances of self-deception, while retaining both the cognitive biasing mechanism and the general doxastic structure of the subject acquiring or maintaining the false-belief that-p. However, the functional objective of the cognitive bias is re-conceptualized in terms of mental control and automaticity.

Mental control refers to an agent's ability to intentionally regulate and direct thoughts, beliefs, behavior, and emotion when there is no automatic response to a stimulus. It is initiated by a conscious and intentional goal-directed process (which is temporally very brief), and then fulfilled and maintained by means of automatic processing. There are five stages of mental control. The initial stage is the emergence of consciousness in the context of

a situation for which there is no previously determined automatic response, such as new, unfamiliar, or emotionally intense circumstances. This is followed by conscious acknowledgement, which is the initiation of conscious processing. The third stage is intention formation. The intention to control is manifested via the formation of a corresponding memory representation, also known as a "prospective memory" or "intentional plan," which remains in a state of heightened activation to facilitate its quick accessibility. During the automatic support stage, the conscious intention has been formed and its corresponding 'prospective memory' is highly accessible, and can be activated and fulfilled unconsciously and automatically by an appropriate environmental stimulus or cue[6]. The final stage is appraisal, in which the automatic or conscious processes evaluate the progress or success of the control process in realizing the intended goal.

In this alternative picture, the initial stages of the process of self-deception are framed in terms of mental control, i.e., a conscious and intentional goal-directed process. Specifically, the agent's conscious intention to self-deceive plays a crucial but transitory role in initial stages of conscious acknowledgement, in which the subject is aware of his intention to believe that-p. The subsequent intention formation stage triggers the subject's intentional cognitive bias towards the activation of the mental schema that-p in the context of any experience or evidence, positive or negative, relating to that belief. The subject forms a prospective memory for this intention, leaving the schema for the belief that-p in a state of heightened excitation and accessibility and priming it for recurrence. Due to the schema's repeated activation in the context of relevant cues, it becomes strengthened and thus more automatic. Eventually the schema for the belief that-p is chronically accessible and activated automatically in the context of any relevant cues, and in the absence of conscious acknowledgment or intention.

A synopsis of the alternative picture of self-deception that I propose follows:

> The agent forms an intention to believe that-p (a false belief), which is not epistemically warranted. The agent's intention triggers a cognitive bias toward taking an experience as that-p priming in the presence of any contextual cues related to or evidence in support of *either* the belief that-p *or* the belief ¬p. Repeated activation results in the agent's automatic, non-intentional activation of the schema for the belief that-p in the context of any relevant contextual cues.

Returning to the cases of Sam and Tom can provide an illustration of how this process might be executed. Initially, Sam's belief that his wife is faithful is supported by all available evidence. However, the balance of relevant evidence shifts from supporting the belief that she is faithful (that-p), to supporting the belief that she is not (¬p). At this stage Sam engages in mental

control to maintain the (now false) belief that his wife is faithful. This triggers a cognitive bias towards taking all contextual clues regarding Sally's faithfulness (or lack thereof) as activating the belief that-p, and forms a prospective memory to that effect. He is not required to assess evidence or potential implications, because he takes any relevant experience as that-p priming. Upon repeated activations, the schema is primed and chronically accessible enough to be activated automatically in all relevant contexts, without any intentional initiation on Sam's part. Though Sam is no longer intending to deceive himself, he has entered into self-deception. In Tom's case, the emotional trauma over his father's death has resulted in his intention to form the belief that his father was loving and attentive (that-p), despite strong evidence to the contrary (supporting the belief ¬p). This intention triggers a cognitive bias, like Sam's, toward taking any experience involving a memory of or evidence concerning his father, whether positive or negative, as priming the belief that-p. Eventually the belief that-p schema is sufficiently strengthened and chronically accessible such that it can be activated automatically by any encounter with contextual cues concerning his father. At this stage Tom is no longer intending to deceive himself, but has entered a state of self-deception with respect to his belief that his father was kind and loving.

The process described above differs from Mele's in some crucial ways in attempting to address some of the potential weaknesses in his approach. Introducing the agent's intention to acquire or maintain a belief (which happens to be self-deceptive) in the initial stages of the process of self-deception emphatically differentiates the motivated, biased belief formation of the self-deceiver from the motivated, biased belief formation of the wishful thinker. Furthermore, ascribing to the self-deceiver a motivated intention to believe that-p (in however a transient or precursory role) provides an impetus for the agent to resolutely cling to that belief in the face of all contradictory indications, which seems lacking in Mele's employment of desire for the same purposes.

The shift in the functional objective of the cognitive bias addresses the sorting problem raised by Mele's account. In the alternative picture, the agent's bias causes him to take a relevant experience as priming the mental schema for the belief that-p, as opposed to his encountering or interpreting evidence as supporting that belief. The agent eventually conditions himself to take all experiences with contextual cues related to p, regardless of whether those cues support the belief that-p or the belief ¬p, as priming the activation of the schema that-p. There is therefore no need for the agent to assess, interpret, or filter available evidence to facilitate attentional focus on desirable information, a process which might have the concomitant effect of actually reinforcing the unwanted belief ¬p and priming it for recurrence. Because the evidence is not subject to critical evaluation on any level, the belief that-p is primed regardless of whether it is information in support of the belief that-p or the belief ¬p. Thus, the sorting problem is averted.

VIII. The Role of Self in Self-Deception

A final issue to be discussed is the question of how the acquisition or maintenance of beliefs related to the self might differ from that of ordinary beliefs. Beliefs about the self seem not to be subject to the same constraints of rational and critical scrutiny as other beliefs. It could be that encounters with evidence are approached in a fundamentally different way when they are relevant to the beliefs that shape our core conceptions of ourselves, our close emotional relationships, self-assessments of professional success, and so on. This is reflected in the 'garden-variety' cases of self-deception which are typically cited. It seems that beliefs about the self are much more conducive to being acquired or maintained self-deceptively.

It could be that the reason people tend to be self-deceived with respect to core beliefs related to the self is that those beliefs are more easily acquired and maintained, and relevant available evidence is not subject to the normal rigorous evaluative process of assessment, interpretation, and inference. Such beliefs might be so deeply entrenched that it is possible to consistently reinforce them, even with a scarcity of supporting evidence. With respect to self-deception, this would seem to support a move away from the notion of a specialized cognitive bias in evidence filtering and selective attending and toward a more global type of intentional belief priming, as I have suggested above.

IX. Conclusion

Mele argues that cases of intentional self-deception are possible but not typical, and that in garden-variety cases of self-deception the agent has no intention to self-deceive. I have suggested that there are two possible weaknesses in his account: that there is a 'sorting problem' with respect to the function of the cognitive biases he proposes, and that his construal of garden-variety cases of self-deception is in fact more characteristic of wishful thinking than self-deception. In response, I have presented an alternative view of the phenomenon which retains the belief-structure of Mele's account but shifts the location and functional objective of the cognitive biasing mechanism, and have argued that an initial intention to self-deceive should be attributed to the agent in paradigm cases of self-deception.

UCD School of Philosophy
University College Dublin
Newman Building,
Belfield, Dublin 4, Ireland
E-mail: anna.nicholson@ucdconnect.ie

NOTES

[1] There are in fact two versions of the doxastic paradox: weak (sB that-p & ¬p) and strong (sB that-p & ¬[sB that-p]. As the latter is a straightforward contradiction, most accounts of self-deception focus on the former version. However, the extreme divisionist accounts of King-Farlow (1963) and Rorty (1972) address the 'strong' doxastic paradox.

[2] McLaughlin refers to such cases as intentional self-induced deception.

[3] He stresses that the requirement that p be false is of no importance to the dynamics of self-deception, but included to maintain the correspondence with other-deception, which requires that a person is deceived in believing that-p only if p is false. This is a contentious point — it could also be argued that p's falsity should not be a requirement for other-deception.

[4] I am treating 'wishful thinking' as identical to 'wishful believing' for the present purpose.

[5] Priming or 'temporary accessibility' refers to the recent activation of a schema, which leaves it in a state of heightened excitation, and thus highly accessible in the short term. 'Chronic accessibility' refers to a schema that has been primed or activated so extensively that it remains in a constant state of heightened excitation, and is thus highly accessible in the long term [Bargh (1997)].

[6] Marsh, Hicks & Bink (1998) found evidence that the prospective memory remains in a state of heightened activation until the realization of their associated intention.

REFERENCES

AUDI, R. (1982), "Self-Deception, Action and Will", in *Erkenntnis*, vol. 18, pp. 133-158.

BARGH, J. (1997), "The Automaticity of Everyday Life", in Robert S. Wyer (ed.), *Advances in Social Cognition,* vol. X, New Jersey, Lawrence Erlbaum Associates, Publishers.

CANFIELD, J. & GUSTAVSON, D. (1962), "Self-Deception", in *Analysis*, 23, pp. 2-36.

DAVIDSON, D. (1985) "Deception and Division", originally published in: E. LePore & B. McLaughlin (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford, Blackwell.

HAMLYN, D.W. (1971), "Self-Deception", in *The Aristotelian Society: Supplementary Volume*, vol. 45, pp. 45-60.

JOHNSTON, M. (1988), "Self-Deception and the Nature of Mind." in B.P. McLaughlin and A.O. Rorty (eds.), in *Perspectives on Self-Deception*, Berkeley, CA, University of California.

KING-FARLOW, J. (1963), "Self-Deceivers and Sartrian Seducers", in *Analysis*, vol. 23, pp. 131-136.

LAZAR, A. (1997), "Self-deception and the desire to believe", in *Behavioral and Brain Sciences*, vol. 20, pp. 119-120.

— (1999), "Deceiving Oneself or Self-Deceived?", in *Mind*, vol. 108, pp. 263-290.

MARSH, R.L., HICKS, J.L. & BINK, M.L. (1998), "The Activation of completed, uncompleted, and partially completed intentions", in *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 24, pp. 350-361.

MARSH, R.L., COOK, G.I., MEEKS, J.T., CLARK-FOOS, A., & HICKS, J.L. (forthcoming), "Memory for intention-related material presented in a to-be-ignored channel", in *Memory and Cognition*.

MCLAUGHLIN, B.P. (1988), "Exploring the Possibility of Self-Deception in Belief", in B.P. McLaughlin and A.O. Rorty (eds.), *Perspectives on Self-Deception*, Berkeley, CA, University of California.

MELE, A (1987a), *Irrationality*, Oxford, Oxford University Press

— (1987b), "Recent Work on Self-Deception", in *American Philosophical Quarterly*, vol. 24, pp. 1-17.

— (2003), "Emotion and Desire in Self-Deception", in A. Hatzimoysis (ed.), *Philosophy and the Emotions,* Cambridge, Cambridge University Press, pp. 163-179.

— (1997), "Real Self-Deception", in *Behavioral and Brain Sciences*, vol. 20, pp. 91-102.

PEARS, D. (1985), "The Goals and Strategies of Self- Deception", in J. Elster (ed.), *The Multiple Self*, Cambridge, Cambridge University Press.

RORTY, A.O. (1972), "Belief and Self- Deception", in *Inquiry*, vol. 15, pp. 387-410.

— (1980), "Self-Deception, Akrasia and Irrationality", in *Social Sciences Information*, vol. 19, pp. 905-922.

— (1988), "The Deceptive Self: Liars, Layers, and Lairs", in B.P. McLaughlin and A.O. Rorty (eds.), *Perspectives on Self-Deception*, Berkeley, CA, University of California.

WEGNER, D.M. (1994), "Ironic Processes of Mental Control", in *Psychological Review*, vol. 101, pp. 34-52.