

Nondoxasticism about Self-Deception

Sophie ARCHER[†]

ABSTRACT

The philosophical difficulties presented by self-deception are vexed and multifaceted. One such difficulty is what I call the ‘doxastic problem’ of self-deception. Solving the doxastic problem involves determining whether someone in a state of self-deception that $\sim p$ both believes that p and believes that $\sim p$, simply holds one or the other belief, or, as I will argue, holds neither. This final option, which has been almost entirely overlooked to-date, is what I call ‘nondoxasticism’ about self-deception. In this article, I present a negative case for nondoxasticism according to which, in the paradigm case of self-deception, there is no explanatory need to attribute the self-deceived person either their undesired belief that p , or their desired belief that $\sim p$. Folk psychology is replete with concepts other than belief, and if we bear this in mind, it becomes clear that the explanatory roles for which the self-deceived person’s purported beliefs have traditionally been enlisted can be comfortably filled without recourse to belief.

1. Introduction

The phenomenon of self-deception is philosophically puzzling. Consider the case of a man who is self-deceived over his wife’s infidelity: his wife is having an affair, but he is self-deceived that this is not the case. One typical feature of such a situation is that the man will behave oddly. He may say one thing and do another, for example. Let us imagine that on the gentle prompting of his friends he will always fiercely defend his wife, and yet has taken to coughing loudly whenever he is about to enter a room in which he expects his wife to be alone conducting a telephone conversation. One of the perplexities here is to try and specify what this man *believes*. Does he believe that it is not the case that his wife is having an affair, as what he says to his friends seems to indicate? Or, does he believe (perhaps *know*) ‘deep down’ that she is up to something, as his coughing suggests? Furthermore, the name of the phenomenon in question – ‘self-deception’ – perhaps implies that it ought to be modelled on the ordinary case of deception, ‘other-deception’,¹ except that the man is the self who is *both* ‘the deceived’ who believes that it is not the case that his wife is having an affair (his desired belief), *and*, at the

[†] Department of Philosophy, Keble College, Oxford, OX1 3PG; Email: sophie.archer@ucl.ac.uk

¹ Alfred Mele (e.g., 1997, 92) calls this the “lexical” approach as it involves turning first to the word ‘deception’ and attempting to fit self-deception under its banner. Donald Davidson (e.g., 1986) famously takes exactly this approach to trying to understand self-deception.

The copyright line for this article was changed on 22 July 2014 after original online publication.

same time, 'the deceiver' who believes the contrary (his undesired belief).² In this essay, I will address the problem of specifying what one who is self-deceived believes. More formally, I will address what I call the 'doxastic problem' of self-deception, which requires that one specify whether someone in a state of self-deception that $\sim p$,³ both believes that p and believes that $\sim p$, simply holds one or the other belief or, as I will argue, holds neither.

Attempts to respond to the doxastic problem by modelling self-deception on other-deception have been widely rejected. Such attempts face two main difficulties. First, they must explain how it is that the self-deceived person can both believe that p and believe that $\sim p$, and yet fail to put these two beliefs together to form the impossible belief that p and $\sim p$.⁴ Secondly, if other-deception is understood as (at least typically) intentional, they must specify how one can *intentionally* form a belief while possessing the contrary belief.⁵ Davidson (1986) attempts to solve both of these problems by appealing to the idea that the mind is divided. However, it is far from clear that his attempt to do so is successful.⁶

I will not concern myself with entering directly into the debate as to whether or not such an approach can be made to work; rather, my entry point will be the more recent discussion, which has simply assumed for the sake of argument that the self-deceived person does not possess *both* purported beliefs. In the more recent debate, there are, on the one hand, theorists who claim that in paradigm cases the self-deceived person holds their undesired belief and not their desired belief (e.g., Robert Audi (1997), Eric Funkhouser (2005) and Tamar Szabó Gendler (2007)), and, on the other hand, there are those who claim that the converse is true: in the paradigm case, the self-deceived person does not hold their undesired belief, but they *do* hold their desired belief (e.g., Annette Barnes (1997) and Alfred Mele, as recently as 2010). In what follows, I will defend an option which all parties seem largely to overlook: I will argue for a position that I will call 'nondoxasticism', according to which – in the paradigm case – the self-deceived person neither believes that p nor believes that $\sim p$.⁷

² There is some debate regarding how precisely to understand other-deception. For an informative summary of the debate, see James Mahon (2008).

³ For simplicity's sake, I adopt the standard terminology throughout in discussing self-deception that $\sim p$, and the belief that $p/\sim p$. This terminology implies that both involve propositions. Strictly, however, I wish to remain neutral here as to whether or not this is the case.

⁴ This belief is impossible, provided the believer does not also believe that there can be true contradictions. Graham Priest (2006) is an example of someone who holds that some contradictions are true, but his view is a minority one and we will set it aside.

⁵ Following Mele (e.g., 1997, 92), these difficulties have become known as the 'static' and 'dynamic' puzzles respectively.

⁶ For a convincing criticism of Davidson's position, see David Pears (1984) and Sebastian Gardner (1993) for a convincing criticism of all divided mind approaches.

⁷ Eric Funkhouser (2009), Eric Schwitzgebel (2010), and very recently José Porcher (2012) have all pointed to the possibility of denying that the self-deceived person possesses either

My strategy will be to argue that there is no explanatory need to attribute the self-deceived person either belief: the explanatory requirements for which the self-deceived person's putative beliefs have traditionally been enlisted can comfortably be satisfied by appeal to other psychological categories. Now, it is important to note that nondoxasticism will not have been conclusively demonstrated by this negative argument. In order to provide such a conclusive demonstration, one would need to provide a positive argument for preferring the nondoxasticist solution to the other three on the table. I hope to provide such a positive argument in further work but, for now, I am content simply to draw attention to the viability of the much neglected nondoxasticist solution to the doxastic problem, and to emphasize that its viability depends upon recognizing the richness of folk psychology.

I will proceed as follows. I will begin by outlining the four main solutions to the doxastic problem. I will then explain why I will be focusing solely on paradigm cases of self-deception and I will provide a rough-and-ready characterization of what I mean by a 'paradigm' case. I will go on to take each purported belief – the self-deceived person's undesired belief and their desired belief – in turn and consider the explanatory role it is alleged to play in isolation from the other. I will argue that, considered independently, there is no explanatory need to postulate either belief. Of course, my negative conclusion does not follow straightforwardly from this, as there may be some reason(s) to think that the self-deceived person must hold *at least one* of the beliefs in question.⁸ Thus, I will go on to examine how attributing the self-deceived person one belief might be supposed to interact with failing to attribute them the other, as well as some additional reasons one might think that the self-deceived person need hold at least one of the beliefs. I will argue that none of these considerations are convincing. So, I will conclude that, in paradigm cases of self-deception, no explanatory need to attribute the self-deceived person either their undesired or their desired belief has been demonstrated.

2. *The options*

There are four types of response to the doxastic problem:

their undesired or their desired belief. However, although they are thus strictly nondoxasticists according to my terminology, we disagree fundamentally. They think that the self-deceived person is, as Schwitzgebel puts it, "in-between believing" that *p* and lacking this belief and that folk psychology is limited in its capacity to account for self-deception. In what follows, I argue that this is not the case and I eschew the idea that what the self-deceived person believes is indeterminate. See my Edwards (unpublished ms) for a critique of the notion of in-between believing.

⁸ Thanks to Lucy O'Brien for emphasizing the importance of this point.

- (1) To maintain that the self-deceived person both believes that p (their undesired belief) and believes that $\sim p$ (their desired belief). (Davidson (e.g., 1986) is the most famous proponent of this position (at least for self-deception in its 'strongest' sense – see Davidson (1986, 208)).)
- (2) To maintain that the self-deceived person believes that p (their undesired belief) but they do not believe that $\sim p$ (their desired belief). (Proponents include Robert Audi (1997), Eric Funkhouser (2005), Tamar Szabó Gendler (2007).)
- (3) To maintain that the self-deceived person does not believe that p (their undesired belief), but they do believe that $\sim p$ (their desired belief). (Proponents include Annette Barnes (1997), and Alfred Mele, as recently as (2010).⁹)
- (4) To maintain that the self-deceived person neither believes that p (their undesired belief) nor believes that $\sim p$ (their desired belief).

So, if we take the case of the husband who becomes self-deceived that his unfaithful wife is not having an affair, according to option (1), he both believes that she is having an affair and that it is not the case that she is having an affair. According to option (2), he believes that she is having an affair but he fails to bring himself to believe that it is not the case that she is having an affair. According to option (3), he does not believe that she is having an affair, but rather manages to bring himself to believe that it is not the case that she is having an affair. And, finally, according to option (4), he neither believes that she is having an affair nor believes that it is not the case that she is having an affair.

3. Paradigm cases

We have seen that modelling self-deception on other-deception results in affirming option (1). As discussed, this response has its difficulties. Furthermore, it would be simply stipulative to insist that modelling self-deception on other-deception is the only approach to understanding self-deception. If we assume from the outset that self-deception will not count as 'deception' unless it is modelled on other-deception, then our discussion of the doxastic problem will be a very short one.¹⁰ So, how else are we to approach the phenomenon of self-deception in order to

⁹ Mele only ever offers a list of sufficient conditions for entering self-deception, and so, strictly speaking, does not claim that the self-deceived person must possess their desired belief. In fact, in his (2010), Mele admits that it is consistent with someone's being self-deceived that they merely believe that they believe their desired belief (without actually believing it). Nonetheless, his position is best characterized as an example of option (3) as elsewhere he consistently claims that the self-deceived person holds their desired belief (and fails to possess their undesired belief).

¹⁰ Mele (e.g., 1997, 92–93) makes this point.

examine the doxastic problem with fresh eyes? It seems reasonable to maintain that our approach should be to attempt to look directly at particular cases of the phenomenon of self-deception itself, with as few theoretical assumptions as possible.

However, self-deception is attributable to people in a very wide variety of cases. In fact, there is so much disparity in cases of self-deception that a unified account of the phenomenon may be difficult to provide. Does this mean that we are forced to admit that we can only pass comment on a case-by-case basis? I do not think it does. Rather, my strategy will be to isolate 'paradigm' cases of the phenomenon and limit my response to the doxastic problem to range over only these cases. Indeed, there is disagreement in the philosophical literature on self-deception even about how to account for what are *agreed* to be paradigm cases of the phenomenon. Now, although some of this disagreement perhaps may be explained away by the manner in which each party emphasizes different features of the case in question, it seems charitable to assume that if we can abstract from such emphases, we will reach genuine disagreement between authors as to how to construe paradigm instances of the phenomenon. Assuming this to be the case then, I will address the doxastic problem by examining paradigm cases of self-deception (attempting myself to stick to as neutral a presentation of each as possible, of course). So, from now on, unless I specify otherwise, when I refer to 'self-deception' I mean to refer to paradigm cases of the phenomenon.

So, what counts as a 'paradigm' case of self-deception? Following Mele again (e.g., 2001), it is common to distinguish between what he calls 'straight' and 'twisted' versions of the phenomenon. Mele defines 'straight' self-deception as involving someone's becoming self-deceived into believing that something they want to be the case *is* the case. Consider the husband who desires that it is *not* the case that his wife is having an affair becoming self-deceived that this is so. 'Twisted' self-deception, on the other hand, involves someone's becoming self-deceived that something they do *not* want to be the case (and do not also want to be the case) *is* the case. Consider a jealous husband who does not desire to believe that his wife is having an affair, who becomes self-deceived that it *is* the case that she is so-engaged.

A caveat is necessary here. Characterizing precisely *which* desires are involved in both straight and twisted versions of self-deception is a complex matter. For one thing, one might think that 'twisted' cases such as that of the jealous husband are better characterized in terms of his (unconsciously) desiring to believe that his wife *is* having an affair as he thinks (again, presumably unconsciously) that such a belief will help protect him from the harm he would incur if he were duped into believing that this was not so, when she was in fact so-engaged. I will not be interested in twisted cases here, so settling this dispute will not matter for my purposes. Following my predecessors in the debate, I will include only straight

cases under the banner of ‘paradigm’ cases.¹¹ Nonetheless, a second issue presents itself regarding the self-deceived person’s desires even once we have narrowed our focus to straight cases. Mele characterizes the desire of the straight self-deceiver as the desire that $\sim p$ be the case, but I have been talking so far in terms of desired *beliefs*, rather than desired states of affairs. Which of these desires are involved requires an independent discussion in its own right, but I suspect that Sebastian Gardner (1993, 18) is correct when he says that:

It would be wrong . . . to see self-deception as resulting from a preference for trying to solve an internal or psychological problem over its external or real counterpart. Instead, self-deceivers should be seen as mistakenly taking themselves to have solved their real problem in solving their psychological problem; or, put another way, as failing to make a proper distinction between psychological and real problems.

For the sake of simplicity, I will talk throughout of the self-deceived person as desiring to believe that $\sim p$, and desiring not to believe that p .

Limiting myself to straight cases then, I will understand paradigm cases of self-deception roughly along the lines of the kinds of cases Audi (1982, 134) describes here:

the lover who cannot bear the thought that his beloved is drawing away, the alcoholic who cannot admit that he is unable by himself to stop drinking, the terminal patient unable to face his death, and the athlete unable to reconcile himself to his waning powers.

Let us now take each of the purported beliefs of the self-deceived person in such cases in turn to begin to examine whether or not it is explanatorily necessary to attribute it to them. First: the undesired belief.

4. *The undesired belief*

Various features of the self-deceived person’s situation have been drawn attention to in order to justify attributing them their undesired belief. I will now consider six of the most common ones and argue that none of them warrants the attribution of this belief.

I will begin by considering the idea that the self-deceived person’s *behaviour* justifies attributing them their undesired belief. Audi (1982) (a proponent of (2)) argues that this is the case: the self-deceived person’s behaviour supports the idea that they *unconsciously* hold their undesired belief. For argumentative purposes, we will understand ‘unconsciously’ holding a belief here as Audi does, to mean holding a belief that does not just *happen* to be outside of consciousness at a certain time, but rather as holding a belief for which there is some psychological

¹¹ In particular, I have in mind Audi and Mele.

barrier to one's consciously countenancing it. Audi (Ib., 137) conceives of the self-deceived person's unconscious belief along the following lines:

unconscious beliefs are very much like conscious ones, apart from two major differences: (i) if they manifest themselves in *S*'s consciousness, he is very unlikely, without special self-scrutiny or prompting from someone else, to attribute these manifestations to them; and (ii), he is, with the same exceptions, very unlikely to explain actions of his which are due to them, *as* due to them. But – and this is the important point here – they do tend to manifest themselves in consciousness and behavior, and in essentially the same way as conscious beliefs, though usually less frequently . . .

Roughly, we might say that *S*'s belief that *p* is unconscious if and only if (1) he does not know or believe he has it, and (2) he cannot come to know or believe he has it without either outside help . . . or some special self-scrutiny.

Let us now consider two of Audi's examples of paradigm cases of self-deception that he claims testify to the fact that the self-deceived person unconsciously holds their undesired belief in this sense. Against Audi, I will argue that we do not need to resort to the claim that the self-deceived person holds their undesired belief in such cases in order to explain those aspects of their behaviour that superficially point to this conclusion.

First, there is Audi's case of Ann. Audi (1982, 134) asks us to imagine that:

Ann is dying of cancer and is aware of many facts, such as her long, steady decline, pointing to this outcome, though no one has told her that her case is terminal and she has avoided letting her doctor give her a prognosis. Suppose further that she talks of recovery and discusses her various plans for the long future . . . the facts pointing to her death are not unmistakably prominent and her talk of recovery is apparently sincere . . . she has better than average medical knowledge . . . [but] (among other things) . . . her talk of recovery lacks full conviction (or exhibits too much apparent conviction), and . . . is often followed by depression or anxiety.¹²

He describes how Ann "asks about funeral arrangements 'just in case'" and "rewrites her will" etc. (Ib., 139). Such behaviour warrants the attribution to Ann of the unconscious belief that she will soon die of cancer, according to Audi. But it seems that there are other ways of accounting for this behaviour on Ann's part that do not involve attributing her this unconscious belief. Mele (2009, 273) (a proponent of (3)) suggests that we say that she *consciously* believes "that there is a significant chance" that she will soon die of the disease, which, together with her relevant desires, explains why she makes out her will etc. However, this explanation seems untrue to the phenomenology of self-deception. It seems that part of the experience of being self-deceived (at least in the paradigm case) is the avoidance of confronting the possibility of the undesired situation too explicitly, so to

¹² A delusion known as *anosognosia* is characterized by the denial of illness. A patient suffering from anosognosia may deny that a limb they are unable to move at all is paralysed, for example. Ann's case is notably distinct from such cases in its severity and, as such, it seems that Audi is entitled to describe her case as one of self-deception as opposed to delusion.

attribute the self-deceived person the conscious belief that there is a chance that this situation obtains rings untrue. Rather, in Ann's case, a ready explanation is that Ann is *anxious* about the state of her health and *suspects* that things are worse than she is currently aware, and that she may find out that something she *fears* is the case by probing the matter too much with her doctors etc., so she avoids confronting the matter in the manner in which she would need to in order to form the true belief that she is terminally ill. Nonetheless, her anxiety and suspicions are enough to prompt the actions on Ann's part that seem to suggest that she holds her undesired belief.

Mele (e.g., 1997, 96) also thinks that a suspicion that *p* can often carry the majority of the burden of explaining the self-deceived person's behaviour that appears to point to their holding their undesired belief. However, one might object to the strategy of appealing to a suspicion here for several reasons: I will isolate two main ones. First, one may insist that a suspicion that *p* simply *is* or entails a low *degree* of belief that *p*. Up until this point, I have been simplifying matters in so far as I have only been discussing 'outright' beliefs, but it is sometimes thought that it can be helpful to talk in terms of 'degrees of belief' in addition to, or instead of, outright beliefs.

So, perhaps talking about a 'suspicion' is just a colloquial way of pointing to a low degree of belief. In which case, I have not *really* demonstrated that the self-deceived person's behaviour presents no explanatory need to attribute them their undesired belief in any interesting sense: they *do* possess their undesired belief – to a low degree.

Tackling this line of objection will involve a brief discussion of the general difficulty of making sense of the idea of a degree of belief. In particular, it seems that there is a fundamental tension between the notion of a degree of belief and that of an outright belief, which suggests that the former concept cannot simply be added to a conceptual scheme containing the latter, and neither can it simply be substituted in to replace it. Believing something to be the case, very roughly, involves taking it to be true, in a certain manner.¹³ What, then, is believing something to a certain degree? It cannot be to take that thing to be true *to that degree* as whether or not something is true is a binary matter (in classical logic, at least): the Law of Excluded Middle states that *p* is either true or false; there is no third option. So, perhaps we ought to understand what a degree of belief is in something like the following way. Once one possesses a degree of belief that *p* over a certain threshold (perhaps 0.5), one counts as outright believing that *p*. Below this threshold, one does not. But, if this is the case, and we are to understand

¹³ Bernard Williams (1973) famously claims this 'manner' of taking to true to be one which 'aims at truth'. See David Velleman (2000), Nishi Shah (2003), and Shah and Velleman (2005) for classic recent discussion of this issue.

a suspicion as a *low* degree of belief, the self-deceived person's suspicion that *p* ought not to be understood as tantamount to their believing that *p*.

But what if one were to insist that a low degree of belief simply involves placing the uncertainty that we are trying to express in the *content* of the belief as follows: 'I believe that it is quite likely that *p*'? Indeed, Kevin Lynch (2012, p. 438) thinks that degrees of belief talk can be understood as a formalization of:

everyday locutions as when we claim to be or feel fully convinced or certain, very convinced, fairly convinced, not very convinced, not at all convinced, etc., that *p* (where we sometimes use 'confident' or 'sure' instead of 'convinced').

And, if this is the case, perhaps a suspicion is, or entails, a 'degree of belief' in this sense. Here I think the right thing to say is that we ought to accept that a suspicion that *p* involves (although it cannot be reduced to), at least, something like the belief 'It may be the case that *p*', or perhaps even in some cases, 'It is likely that *p*'. However, if my opponent were to *define* this as involving a degree of belief and hence claim it to be tantamount to admitting that the self-deceived person believes that *p* to some degree, this would clearly be illegitimate. I am not forced to concede that this is to believe that *p* to some degree. Rather, I can insist that it is to possess a distinct belief, with a different content.

Nonetheless, haven't I admitted that the self-deceived person may well believe 'It is likely that *p* is the case'? Even if we refrain from referring to this as a 'degree of belief', have I not come uncomfortably close to admitting that they hold their undesired belief that *p*? And, in what sense is my nondoxasticism about self-deception still a nondoxasticism if it involves such beliefs?

I think that it is indeed the case that the self-deceived person often (perhaps unconsciously, as we have been understanding the term) holds many beliefs with *p* in their content, such as 'I am scared that *p*', 'It might be that *p*', or 'It is likely that *p*'. Nonetheless, I still think that it is significant to point out that there is no explanatory need to attribute them the belief that '*p*', in order to explain their behaviour. My nondoxasticism is a nondoxasticism insofar as it denies that there is any explanatory need to attribute the self-deceived person either of the two main purported beliefs. And, as I have argued, it is not merely a matter of terminology that I avoid doing this as it would be to insist that in having a suspicion that *p*, they believe that *p* to a certain degree, which is tantamount to 'believing' it.

A second line of objection one might take to appealing to a suspicion that *p* on the part of the self-deceived person here is to maintain that to suspect that *p* involves being disposed, *ceteris paribus*, to investigate in an even-handed manner whether or not *p* is in fact the case. But, rather than straightforwardly investigating whether or not their undesired state of affairs obtains, the self-deceived typically do all they can to *avoid* discovering that it does. Does this suggest that a suspicion is not the appropriate mental state to appeal to here? No: even if we grant that

suspecting that *p* does involve, *ceteris paribus*, investigating in an even-handed manner whether or not *p*, in the case of the self-deceived, all else is not equal.¹⁴ The self-deceived person is *anxious* that *p* and they *desire* to believe that $\sim p$ etc. and this explains why their suspicion that *p* does not prompt them genuinely to seek out evidence as to whether or not *p* is the case, but instead disposes them to misinterpret and avoid evidence that *p* and seek only evidence that $\sim p$.

Let us now look to a second case of Audi's (1982, 48) to further examine the behaviour of the self-deceived person. This involves a woman who is self-deceived about the fact that her husband is a "failure". Again, Mele suggests that a belief other than the unconscious belief that her husband is a failure can explain the woman's behaviour that appears to point to her holding this belief. He thinks that the belief that "her husband has failed in many ways" could help explain the woman's negative behaviour towards her husband, or her treating him as she would a failure, without her believing him to *be* a failure (Ib., 161). Again, it seems that the phenomenology of paradigm cases of self-deception precludes us from thinking of the woman as too readily attending to any such belief. However, Mele's suggestion can be taken on board, provided that we recognize that she fails to do so – perhaps it is unconscious in Audi's sense. This belief, combined with a *resentful attitude* towards these failures, is sufficient to explain the woman's negative behaviour towards her husband. Once more, there is no explanatory need to suppose that the self-deceived person unconsciously holds her undesired belief.

In addition to their behaviour, a second consideration that might seem to support the attribution to the self-deceived person of their undesired belief is the idea that their holding such a belief helps *cause* and *sustain* their self-deception. Davidson (e.g., 1986, 208) thinks that this is the case – at least in the 'strongest' cases of self-deception. To examine this idea, let us take the case of the husband who is self-deceived over his wife's infidelity as an example. It might be thought that it is the belief that his wife is having an affair (along with his strong desire to believe that this is not the case, for example) that prompts him to enter into and sustains his self-deception. However, it seems that there is no need to attribute the husband his undesired belief in order to explain the initiation and sustenance of his self-deception. A suspicion that his wife is unfaithful to him on the part of the husband, combined with a strong desire to believe that this is not the case, fill this explanatory role.

¹⁴ One might object that suspecting that *p* does not – even *ceteris paribus* – dispose one to investigate whether *p*: I may suspect that few of my students are yet to start work on an essay that is soon due, but not be disposed to investigate whether this is the case in the slightest. (Thanks to Jonathan Way for this example.) Here, as in all such cases, whether one claims such a disposition to be involved with suspecting that *p* largely depends on when one counts all other things as equal. Am I disinclined to investigate my students because of my depression regarding their work ethic? In which case, perhaps all other things are not equal here.

A third consideration is the self-deceived person's *phenomenology*, to which I have already briefly alluded. William Talbott (1995) considers why this might be thought to point towards the self-deceived person's holding their undesired belief. Talbott (*Ib.*, 44–45) claims that it is the:

almost palpable quality of emotional resistance to doubts about . . . [their desired belief] or evidence against . . . [it] that, I believe, most inclines us to attribute to the self-deceived person at some level a recognition that . . . [their undesired belief is true]. If she does not really recognize that . . . [her undesired belief is true], why does the evidence that . . . [it is] produce such an emotionally charged reaction?

However, as Talbott says, it seems that such reactions can be explained simply by the self-deceived person's awareness that the evidence *could* be taken by them to point towards the truth of the content of their undesired belief, and that if it were, the consequences would be disastrous for them. A suspicion that *p* may even seem to be present in some cases, but, again, it is not explanatorily necessary to attribute them their undesired belief.

A fourth consideration we must consider is what we might call the self-deceived person's 'inputs'. Consider the balding man who is self-deceived about his steadily receding hair-line (this kind of example is employed by Davidson (1986)). We might ask: how can he *fail* to believe that he is balding? After all, he owns a mirror. However, this line of thought presumes that the biasing strategies employed by the self-deceiver, which both control what their inputs are in the first place, and distort the information once it arrives, cannot be substantially successful.¹⁵ Clearly, the uneasiness the balding man feels when his wife motions as to ruffle his hair, and the manner in which he avoids dwelling for longer than a split second on his image in the mirror in the mornings suggest that the biasing strategies are not wholly successful: enough information has got through to make him extremely uncomfortable around the topic of his hair; nonetheless, the suggestion is that the biasing strategies employed by the self-deceiver enable him to avoid forming the belief that he is balding. To stipulate that this cannot be so is to beg the question against one who claims that these strategies *can* be substantially successful.

Fifth, we might draw attention to what ex-self-deceivers typically *say* of their experience of self-deception having come out of their self-deceptive state. When

¹⁵ Mele (e.g., 1997, 94) emphasizes the importance of four such strategies: (1) "Negative Misinterpretation", where one misinterprets as *not* counting (or not counting strongly) against $\sim p$ data that one would easily recognize to count (or count strongly) against $\sim p$ in the absence of self-deception; (2) "Positive Misinterpretation", where one misinterprets as *supporting* $\sim p$ data that one would easily recognize to count against $\sim p$ in the absence of self-deception; (3) "Selective Focusing/Attending", where one both fails to focus attention on evidence that counts against $\sim p$ and to focus instead on evidence suggestive of $\sim p$; and (4) "Selective Evidence-Gathering", where one both overlooks easily obtained evidence for *p* and finds evidence for $\sim p$ that is much less accessible.

they are no longer self-deceived, they may say things such as ‘I knew all along/*deep down* that p was the case’, for example. However, given that the suggestion is that they believed that p unconsciously, in Audi’s sense, it is not obvious that the ex-self-deceiver stands in a privileged relation to their prior self-deceptive state. If they say such things, they would presumably be inferring that these things are the case, just as someone else may infer that the ex-self-deceiver believed their undesired belief, based upon their behaviour. (Barnes (1997, 89) makes a similar point with respect to the self-deceiver’s purported intention.)

A sixth consideration is that we often hold the self-deceived *responsible* for their self-deception – be it rationally, or morally, or sometimes both. Perhaps this would be unwarranted if they did not hold their undesired belief. How can they be criticized if they themselves are not aware of the truth of their situation? However, even granting that we are on occasion correct to censure the self-deceived, this can be explained with or without the idea that they hold their undesired belief. Perhaps they are to be criticized simply for engaging in their biased evidence gathering, for example.

So far, then, I have considered six apparent motivations for attributing the self-deceived person their undesired belief: their behaviour, the causation and sustenance of self-deception, the phenomenology of self-deception, their inputs, what people say of their own past mental states when they are no longer self-deceived, and the fact that we often hold the self-deceived to account over their self-deception. I have argued that none of these considerations demonstrate an explanatory need to suppose that the self-deceived person holds their undesired belief. In particular, I have drawn attention to the fact that folk psychology is sufficiently rich to be able to explain the behaviour and phenomenology of self-deception without appealing to the idea that the self-deceived person unconsciously believes their undesired belief. By way of example, I have suggested that a *suspicion* that p , combined with *anxiety* that p , and a *desire* to believe that $\sim p$ may meet the explanatory charge in some cases.

5. *The desired belief*

I will now turn to the self-deceived person’s desired belief. Again, I will begin by considering whether or not their behaviour warrants the attribution of this belief to them. As with their undesired belief, *prima facie* the self-deceived person behaves as if they believe that $\sim p$. Audi’s Ann, who is dying of cancer, makes plans for the long-term future and talks as if it is not the case that she will soon die etc. However, Audi appeals to his example of Ann in order to argue that the self-deceived person’s behaviour does *not* support the attribution of their desired belief to them. He describes how Ann’s “talk of recovery lacks full conviction (or exhibits too much apparent conviction)” and so does not justify the attribution to

her of the belief that it is not the case that she will soon die of cancer (Audi 1982, 134). Furthermore, according to Audi (Ib., 139), behaviour that seems to evidence the self-deceived person's holding their desired belief could equally be explained in terms of their "wants, needs, etc. that explain why the unconscious belief is unconscious in the first place". This is somewhat suggestive. Is there more that could be said here?

We could begin by looking to Audi's own suggestion that the self-deceived person simply *sincerely avows* that $\sim p$, where sincere avowal is understood by Audi as not entailing belief. Audi thinks that the self-deceived person sincerely avows that $\sim p$ insofar as they lack the belief that they are speaking falsely in so doing, lack any intention to deceive, are not disposed to tell themselves or those they trust that p , and so on. Is the sincerity of one's speech acts secured simply by failing to believe that one is speaking falsely? It seems that Audi is correct that the person who speaks without any belief that what they are saying is false etc. is not straightforwardly *insincere*. However, it also seems that more is required of one who is sincere in what one says: in order to speak sincerely, perhaps one need believe what one says, or perhaps one need only believe that one believes it.

We will not pursue this issue however, as it is not clear that we need think of the self-deceived person's speech acts as sincere in any case: they seem to lack "full conviction (or exhibit . . . too much apparent conviction)" (Ib., 134). Given that this is so, one ready explanation of what the self-deceived person is doing when they assert that $\sim p$, for example, is what Mele (1987, 144) calls "acting as-if". As Mele explains, acting as if one's desired belief were true can help generate 'evidence' in favour of one's desired belief, as one can be swayed by one's own behaviour. It can also "generate 'social' evidence for a favored hypothesis": if the self-deceived person acts as if something is the case, then others may well respond to them as if this is the case, which generates further reassurance from the self-deceived person's point of view (Ib., 158). As Mele (Ib., 157) says, "it is not difficult to see why someone may be motivated to act as if . . . [$\sim p$] by a desire . . . [to believe that $\sim p$]." We can also think of the self-deceived person as engaging in a kind of 'internal' acting as-if, in addition to their external behaviour of this kind. Not only do they tell others that $\sim p$, they continually *tell themselves* this as well.¹⁶ In addition to their *anxiety* that p , their *hope* that $\sim p$, their *desire* to

¹⁶ One might wonder whether we can continue to maintain that the self-deceived person employs such biasing strategies if they fail to attain what these strategies are geared towards attaining: their desired belief. However, it seems that, in general, simply because one fails in one's exploits, this is not a good reason to deny that one's behaviour nonetheless manifests the fact that these *are* one's exploits.

believe that $\sim p$ etc., such internal acting as-if could also help lead the self-deceived person to behave as if $\sim p$.¹⁷

So, there are viable alternatives for explaining those aspects of the self-deceived person's behaviour that appear to point to their holding their desired belief. Given this, let us move on to consider any further reasons one might have for attributing the self-deceived their desired belief. Let us briefly consider the phenomenology of someone in a self-deceptive state once more: any attraction the self-deceived feels when the thought $\sim p$ is entertained is readily explicable in terms of their desire to believe that $\sim p$ and to avoid believing that p . Also, as before, the ex-self-deceived person's subsequent assertions with respect to their past self-deception need be granted no special authority. And fourthly, any responsibility we are likely to attribute to the self-deceived person for their self-deceptive state, even if well-placed, does not require the idea that they believe the false, desired belief. They are (epistemically) rationally – and perhaps sometimes morally – criticizable simply for their biased evidence gathering.

In this section, I have considered four apparent motivations for attributing the self-deceived person their desired belief: this time, I have considered their behaviour, the phenomenology of self-deception, what people say of their own past mental states when they are no longer self-deceived, and the fact that we often hold self-deceived people to account over their self-deception. As with the self-deceived person's undesired belief, I have argued that these considerations demonstrate no explanatory need to attribute the self-deceived their desired belief.

6. Putting the two together

So far then, we have seen that, considered in isolation, no explanatory need has been demonstrated to attribute the self-deceived person either their undesired belief that p , or their desired belief that $\sim p$. Of course, it does not follow

¹⁷ Gendler (2007) makes a similar suggestion with respect to explaining the self-deceived person's behaviour that appears to evidence their believing that $\sim p$. She claims that the self-deceived person does not believe that $\sim p$; rather, they engage in what she calls *imaginative pretense* that $\sim p$, which involves being engaged: 'in some sort of activity that involves . . . thinking about, in more or less detail, what things would be like if . . . the content of . . . [their] imagining were actually the case'.

She also acknowledges that: 'at the same time, if circumstances allow, . . . [they] might reinforce . . . [their] fantasy by *performatively pretending* (in the sense of non-believingly acting as if it were the case) that not-P – speaking to others as if not-P were the case, governing . . . [their] actions as if not-P held' (Ib., 241, my italics).

Her *imaginative pretense* corresponds to my *internal acting as-if* and her *performative pretense to acting as-if*. However, Gendler's position differs from mine in so far as she thinks that in engaging in these forms of pretence that $\sim p$, the self-deceived person nonetheless holds their undesired belief that p .

straightaway from this that in the paradigm case there is no explanatory need to attribute the self-deceived either belief. What if there were some reason(s) to think that they must have at least one of the beliefs in question? We will begin by considering the possibility that attributing the self-deceived one of their beliefs is necessary for withholding the attribution of the other belief.

According to position (2), on which the self-deceived person holds their undesired belief that p , but not their desired belief that $\sim p$, it could be thought that their undesired belief helps to explain why they fail to attain their desired belief. Given the assumption that they do not hold contradictory beliefs, the idea could be that they are frustrated in their attempts to acquire their desired belief by their prior possession of their undesired belief.

On position (3), according to which the self-deceived person does not hold their undesired belief, but does hold their desired belief, the fact that we attribute them their desired belief could be understood as explaining why we fail to attribute them their undesired belief. Again, given the assumption that they do not hold contradictory beliefs, the fact that we are prepared to attribute them their desired belief might be thought to explain why we cannot attribute them their undesired belief.

However, it seems that these explanatory roles can be met without attributing the self-deceived person either belief. Let us take the case of the husband who is self-deceived over his wife's infidelity as an example once more. According to the nondoxasticist, the husband begins to *suspect* that his wife may be having an affair and, immediately, his defences go up. He strongly *desires* to believe that his wife is faithful to him. The combination of this desire and his suspicion explains why he begins to avoid information in favour of his feared conclusion and seek out evidence against it, and distort any evidence he finds. Engaging in such biased evidence gathering, he prevents himself from coming to hold his undesired belief that his wife is having an affair. Even so, there is significant disquiet in his mind regarding the issue: he is *anxious* that it not be the case that his wife is having an affair, he has niggling *doubts* about the information he already has, but he *hopes* that it is not the case that she is unfaithful to him. Nonetheless, his niggling doubts and suspicions prevent him from attaining the belief that it is not the case that his wife is having an affair.

On this construal of the case, the explanation as to why the husband fails to attain his desired belief is that his niggling doubts and suspicions to the contrary prevent him from so doing. We need not stipulate that *only* a belief that p can prevent one from forming the belief that $\sim p$. Similarly, there appears to be no reason to insist that only the possession of his desired belief prevents the husband from failing to hold his undesired belief. The biasing mechanisms described ensure that he fails to hold his undesired belief, whether or not they result in his coming to believe his desired belief.

So, we have seen that attributing the self-deceived person one of the beliefs in question is not necessary for withholding the attribution of the other. But are there any further reasons to think that the self-deceived must hold at least one of the purported beliefs?

First, one might think that if someone feels drawn to both the belief that p and the belief that $\sim p$, yet holds neither, they are in a state of simple ambivalence with respect to p ; they are not self-deceived about p . So, perhaps, in order to distinguish self-deception from ambivalence, we must attribute the self-deceived person either the belief that p or the belief that $\sim p$. However, it seems that there are at least two ways in which the nondoxasticist can distinguish self-deception from simple ambivalence. First, we can draw attention to the fact that self-deception is necessarily motivated: the self-deceived person's desire to believe that $\sim p$, as well as their anxieties surrounding the subject-matter about which they are self-deceived, mark out self-deception from simple ambivalence. Furthermore, we might point to the fact that if one is simply ambivalent about whether p , one will typically *behave* very differently than if one is self-deceived concerning p . For example, if one is merely ambivalent about p , *ceteris paribus*, one will affirm that this is so, whereas in the paradigm case, one in a state of self-deception will vehemently affirm that $\sim p$.

Secondly, one might think that in order to explain why self-deception was ever so-called in the first place, one needs to retain *something* of the model of ordinary deception, even if one is not a slave to it to the extent that one must affirm solution (1) to the doxastic problem. Why is self-deception called 'deception' if the self is neither the deceiver who believes the truth, nor the deceived who believes a falsehood? But the nondoxasticist can maintain that self-deception acquired the name it did because the self-deceived person's behaviour superficially suggests that they hold both their undesired and their desired beliefs (and, of course, because self-deception is apparently intentional), and so *appears* to be both deceiver and deceived. The nondoxasticist does not need to appeal to the self-deceived person's possession of at least one of the beliefs in question to explain why the name given to the phenomenon under discussion is 'self-deception'.

Finally, one might think that we need to attribute to the self-deceived person either their undesired or their desired belief in order to explain the irrationality involved in self-deception. Perhaps the self-deceived person is irrational insofar as either (1) they believe the truth, but act as if it were not so, or (2) because they have come to hold an irrational belief in a falsehood. However, the nondoxasticist can readily explain the irrationality involved in self-deception without appealing to either belief: self-deception is (epistemically) irrational simply insofar as it involves biased evidence gathering, and, in some cases, a failure to believe what the evidence one possesses supports.

7. Conclusion

I have argued for a solution to the doxastic problem of self-deception that I have called ‘nondoxasticism’. According to this position, in the paradigm case of self-deception, someone in a state of self-deception neither holds their undesired belief that p nor their desired belief that $\sim p$.

I presented a negative case in support of this claim. I began by taking each of the self-deceived person’s purported beliefs in turn and argued that, considered in isolation, there is no explanatory need to attribute either of them. I then argued that there is no need to say that, in the paradigm case, the self-deceived person need hold at least one of the beliefs in question. I began by showing that the attribution of one belief is not necessary to explain why we withhold the attribution of the other. The self-deceived person fails to attain their desired belief because of their giggling doubts and suspicions otherwise. They fail to attain their undesired belief because of the success of the biasing strategies they begin to employ as soon as they suspect that things are not how they would like them to be. Furthermore, I argued that the distinction between self-deception and simple ambivalence is easy for the nondoxasticist to draw – either in terms of the self-deceived person’s motivation, or their behaviour. They can explain why self-deception was so-called in terms of the surface appearances of the phenomenon. And, finally, the nondoxasticist can account for the irrationality of self-deception without appealing to the self-deceived person’s beliefs, by pointing to the fact that the self-deceived person is motivated to avoid one belief and attain another, irrespective of their truth-values.

In arguing for this position, I have sought to align myself with the idea that ordinary, folk psychology affords a rich array of potentially explanatory concepts, which stretch beyond the two old favourites – belief and desire. I propose that we consider the self-deceiver’s hopes, suspicions, doubts, and anxieties, to name a few, as opposed to thinking of the majority of their behaviour as revelatory solely of belief-desire pairs.*

REFERENCES

- AUDI, R. 1982, “Self-Deception, Action, and Will”, *Erkenntnis* **18**, 2, pp. 133–158.
 AUDI, R. 1997, “Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele”, *Behavioural and Brain Sciences*, **20**, p. 104.
 BARNES, A. 1997, *Seeing Through Self-Deception*, Cambridge: Cambridge University Press.

* For extremely valuable discussion of the present material, I would like to thank Robert Audi, Tim Crane, Max Edwards, Craig French, Ciaran Kamp, Rory Madden, Alfred Mele, Lucy O’Brien, Anthony Price, and Paul Snowdon. For greatly instructive feedback, I am also grateful to two anonymous referees for the present journal, as well as several audiences in London and one in Oxford.

- DAVIDSON, D. 1986, "Deception and Division", in: M. Cavell, ed., *Problems of Rationality*, Oxford: Clarendon Press, pp. 199–212.
- EDWARDS, S. unpublished manuscript, "Why 'Believes' is Not a Vague Predicate".
- FUNKHOUSER, E. 2005, "Do the Self-Deceived Get What They Want?", *Pacific Philosophical Quarterly*, **86**, 3, pp. 295–312.
- FUNKHOUSER, E. 2009, "Self-Deception and the Limits of Folk Psychology", *Social Theory and Practice*, **35**, 1, pp. 1–13.
- GARDNER, S. 1993, *Irrationality and the Philosophy of Psychoanalysis*, Cambridge: Cambridge University Press.
- GENDLER, T. 2007, "Self-Deception as Pretense", *Philosophical Perspectives*, **21**, 1, pp. 231–258.
- LYNCH, K. 2012, "On the 'Tension' Inherent in Self-Deception", *Philosophical Psychology*, **25**, 3, pp. 433–450.
- MAHON, J. 2008, "The Definition of Lying and Deception", in: E. Zalta, ed., *The Stanford Encyclopedia of Philosophy*, online at: <http://plato.stanford.edu/archives/fall2008/entries/lying-definition/>
- MELE, A. 1987, *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*, Oxford: Oxford University Press.
- MELE, A. 1997, "Real Self-Deception", *Behavioural and Brain Sciences*, **20**, pp. 91–136.
- MELE, A. 2001, *Self-Deception Unmasked*, Oxford: Princeton University Press.
- MELE, A. 2009, "Have I Unmasked Self-Deception or Am I Self-Deceived?", in: C. Martin, ed., *The Philosophy of Deception*, Oxford: Oxford University Press, pp. 260–276.
- MELE, A. 2010, "Approaching Self-Deception: How Robert Audi and I Part Company", *Consciousness and Cognition*, **19**, 3, pp. 745–750.
- PEARS, D. 1984, *Motivated Irrationality*, Oxford: Clarendon Press.
- PORCHER, J. 2012, "Against the Deflationary Account of Self-Deception", *Humana Mente*, **20**, pp. 67–84.
- PRIEST, G. 2006, *In Contradiction: A Study of the Transconsistent*, Oxford: Clarendon Press.
- SCHWITZGEBEL, E. 2010, "Acting Contrary to One's Professed Beliefs, or the Gulf Between Occurrent Judgment and Dispositional Belief", *Pacific Philosophical Quarterly*, **91**, pp. 531–553.
- SHAH, N. 2003, "How Truth Governs Belief", *The Philosophical Review*, **112**, 4, pp. 447–482.
- SHAH, N. and VELLEMAN, D. 2005, "Doxastic Deliberation", *The Philosophical Review*, **114**, 4, pp. 497–534.
- TALBOTT, W. 1995, "Intentional Self-Deception in a Single Coherent Self", *Philosophy and Phenomenological Research*, **55**, 1, pp. 27–74.
- VELLEMAN, D. 2000, *The Possibility of Practical Reason*, Oxford: Clarendon Press.
- WILLIAMS, B. 1973, "Deciding to Believe", in: B. Williams, ed., *Problems of the Self*, Cambridge: Cambridge University Press, pp. 136–151.