Self-Deception and Shifts of Attention

Kevin Lynch

Abstract

A prevalent assumption among philosophers who believe that people can intentionally deceive themselves (intentionalists) is that they accomplish this by controlling what evidence they attend to. This article is concerned primarily with the evaluation of this claim, which we may call 'attentionalism'. According to attentionalism, when one justifiably believes/suspects that not-*p* but wishes to make oneself believe that *p*, one may do this by shifting attention away from the considerations supportive of the belief that not-*p* and onto considerations supportive of the belief that *p*. The details of this theory are elaborated, its theoretical importance is pointed out, and it is argued that the strategy is supposed to work by leading to the repression of one's knowledge of the unwelcome considerations. However, I then show that the assumption that this is possible is opposed by the balance of a relevant body of empirical research, namely, the thought-suppression literature, and so intentionalism about self-deception cannot find vindication in the attentional theory.

1. Introduction

According to a common assumption in the philosophical literature about how self-deception gets accomplished, subjects deceive themselves into believing something by the control of attention. Robert Lockie, recognizing the importance of such accounts, has recently labelled them 'attentional accounts' of self-deception or simply 'attentionalism'. According to this, self-deception is 'a phenomenon of attention: one attends to data conducive to the false conclusion, [*p*], and does not attend [that is, deliberately avoids attending] to data conducive to [not-*p*]. In this way one deceives oneself into believing [that *p*]' (2003, 131). This article is concerned with examining this view.

'Attentionalism' will be defined more precisely as we go on, but let me first say that the examination of this view gains particular significance from its close relation to the intentionalist theory of self-deception. Although the term 'intentionalism' can mean different things in debates over self-deception, for present purposes I define it as the view that intentionally deceiving oneself is something that people are *ordinarily able to do* (thus an intentionalist need not hold that self-deception necessarily, or even normally, involves intentionally deceiving oneself). By 'ordinarily' I mean to exclude cases where people might have *special means* available for the purpose, brain-altering technology etc., and so we are only concerned with whether people have the power to intentionally deceive themselves using the ordinary resources of their own mind and body.[1] Moreover, the sense of 'deceives oneself' in use here is what has been called the 'strong' sense (Pugmire 1969), a sense which philosophers have traditionally arrived at from taking self-deception to be the strict intrapersonal analogue of interpersonal deception. In this sense, deceiving oneself means *making oneself acquire a belief which, before trying to do this, one knew to be false or at least unwarranted*, so intentionalism is the view that people do this, or can do this, intentionally.[2]

Now if it is possible for people to do this, then there must be a *feasible strategy* by which they can (excluding the use of special means), and when we look to what intentionalists say about such strategies, we mostly see them mentioning ones involving the shifting of attention between evidence. So attentionalism, we could say, gets adopted by intentionalists as their favoured *strategy theory*, that is, their favoured theory on how people manage to successfully deceive themselves intentionally. Thus, attentionalism would seem to be a major pillar of support for intentionalism and is worth investigating for that reason. I will say a bit about why intentionalists might favour this strategy theory later.

Philosophers who say that people can deceive themselves by the control of attention

---

[1] I am thus following Barnes who says 'I am concerned with what must be realistically done [to intentionally deceive oneself]. Taking a pill, hypnosis, etc. are not realistic options here' (1997, 27). Presumably, these possibilities are not relevant for Barnes because they would not vindicate the idea that intentional self-deception is the common phenomenon some philosophers take it to be.

[2] Some hold that other cases not involving false/unwarranted belief would count as intentional self-deception too. Martin, for instance, thinks that the 'intentional evasion of unpleasant topics and truths' which one knows about (1997, 122) is a kind of self-deception. We will focus on the 'strong' notion of self-deception here.

('attentionalists'[3]) are more often than not found stating this view in passing, or in a casual, cursory fashion, and the question of whether it is true needs to be examined more carefully. In what follows, I will explicate the attentional theory and show how the attentional strategy is supposed to work. I will argue that it presupposes that by shifting our attention off certain thoughts, we can make ourselves forget what we know. But in assuming this, attentionalism runs counter to the balance of the results of some relevant empirical studies, namely, the thought-suppression literature. In light of these studies, self-deception on the attentionalist account does not seem possible, and so intentionalists will have more work to do to convince us that people really can intentionally deceive themselves in the normal course of events.

2. The attentional account of intentional self-deception

An initial formulation of how intentionalists see self-deception as being accomplished through the attentional strategy may now be given. Note that in all the relevant cases, the subject first encounters unwelcome evidence, favouring not-*p*. Stronger or weaker cases will then be possible depending on the evidence's strength. Thus the evidence might be such as to make the subject justifiably believe that not-*p* or at least justifiably *suspect* that this *may* be true. We will focus on the stronger case for simplicity's sake, and because this is the sort of case most commonly discussed by attentionalists (but to consider the weaker case, simply substitute relevant instances of 'believes' with 'suspects'). Moreover, I believe we are not being unfair in doing this since the problems we will encounter with attentionlism would affect the stronger and weaker cases equally (in anticipation, this problem concerns the difficulty one would have in effectively suppressing knowledge of evidence, which is a crucial element of both the strong and weak cases). With that said, how intentional self-deception happens on the attentionalist account is as follows:

---

[3] One could also hold that one could *un*intentionally deceive oneself into believing something by shifting attention in these ways, and indeed, Mele describes such a case (1997, 98). Thus I do not take attentionalism to be essentially tied to intentionalism. We need not go into this further complication however, and we will focus here just on the use intentionalists make of this idea.

A person who after encountering unwelcome considerations (call them 'C^-') and forming the warranted, true, and unwelcome belief on their basis that not-p, but who wishes to believe that p (to avoid the distress of knowing that not-p), by averting his attention from C^-, and by finding and attending to more welcome considerations (C^+) supportive of the contrary proposition p, may cause himself to lose the belief that not-p and acquire the belief that p.

Note that the averting of attention here does not only imply physical acts such as averting one's eyes, but also averting one's *thoughts*, that is, the deliberate avoidance of certain thoughts or the ridding of thoughts from the mind (what psychologists call 'thought-suppression'), as well as selective focusing of attention onto more welcome considerations. For physically avoiding unwelcome evidence would not suffice for the task of self-deception so long as one remembered or remained mindful of the existence and import of that evidence.[4]

An illustration of this phenomenon may be useful at this point, and I hope that the following case captures the kind of thing attentionalists have in mind. When Gloria started dating Tony, she was quickly swept off her feet. He is handsome, charismatic, generous, makes her feel special, and she is impressed with the fact that he seems to have so many friends and seems so respected when they go out around town. He seems to have money too, though he does not talk much about his job in his Uncle's distribution business, for which he keeps odd hours. Their relationship moves fast, and when he asks her to marry him, she agrees. Then, one day soon after their unforgettable wedding, Gloria reads a headline in the newspaper linking Tony's Uncle with organized crime rackets. She ponders: 'could Tony be in the Mob?', and suddenly all these strange details and puzzling incidents about him, for which he always had a separate explanation, fit together like jigsaw pieces (in fact, Tony lives by a strict code prohibiting wives and girlfriends from being privy to 'the business'). Horrified, she immediately shuts the paper, casts out of her mind all thoughts of the article, and reminds herself of Tony's better qualities. By

---

[4] For what follows, it will be useful to bear in mind what suppression actually involves. Daniel Wegner – one of the most prominent researchers on this issue – says that when people try to take their mind off something, they do so by putting it onto something else (Wegner 1994, 12, 60). They turn their thoughts or attention to a 'distracter' which may absorb their attention and precipitate their forgetting. It might also involve simply refraining from elaborating on a thought.

these means she protects, for a while at least, her belief that she has married a decent man (for a similar illustration given by an attentionalist, see Perring 1997).

Next, let us look at some textual evidence showing how frequently such attentional shifts have been implicated in self-deception. In Davidson's opinion, '[t]he action involved [in self-deception] may be no more than an intentional directing of attention away from the evidence in favour of *p*; or it may involve the active search for evidence against *p*' ([1986] 2004, 208). 'An important way to become self-deceived' Perring writes, 'is by intentionally ignoring and avoiding the contemplation of evidence one has for an upsetting conclusion, knowing full well that one is giving priority to one's present peace of mind over the search for truth' (1997, 123). For Whisner, one can lose a warranted belief 'either by shifting attention from a focused attention on the "object" [i.e. the supporting evidence] or by refusing to focus attention on the "object"' (1998, 204). Hamlyn agrees that a man 'can concentrate his attention on certain things to the exclusion of others' and thereby deceive himself (1971, 58), while for Audi, an unwelcome belief that not-*p* can be made 'inaccessible', and '[p]resumably, the process often begins with *S*'s consciously putting the evidence against *p* out of mind, attending eagerly to any evidence there is for *p*, and the like' (1976, 383). And more recently, Van Leeuwen has stated that 'discomfort is there especially when my attention is on the evidence that ~*p*, but abates when my attention shifts–especially when it shifts to evidence that *p* . . . Self-deception ensues, for the focus of attention on the evidence that *p* gives rise to the *p*-belief' (2008, 198–199). Again, we are calling such theorists 'attentionalists'.

Most of these philosophers accept that one may *intentionally* deceive oneself, in the mentioned sense, through such attentional manoeuvres,[5] though some of them (Audi and Whisner, for instance) think that, at least more commonly, such self-deceivers make themselves lose an unwelcome belief, without gaining the contrary welcome one.[6] In fact, when intentionalists discuss what strategies people use to intentionally deceive themselves, they typically mention such behaviour. It is also noteworthy that these philosophers make no special effort to insist that these acts of shifting attention are carried out *unconsciously*, with the subject being *unaware* of what he is doing or why he is doing it, and some are even explicit in saying

---

[5] E.g., see Audi (1982, 152–153), Hamlyn (1971, 46), Perring (1997, 123–124), and Van Leeuwen (2008, 202).

[6] Some might argue that these cases would be insufficient for being deceived since they do not result in the subject having a mistaken belief. They may simply be cases of the intentional repression of belief.

that self-deceivers act like so knowing full well what they are trying to do (see the Perring and Audi quotes above). I believe it is a strength of the attentionalist version of intentionalism that it does not need to rely on the idea that the self-deceiving acts would have to be done unconsciously. For it is commonly held that *doing something intentionally entails doing it knowingly*. Call this the *knowledge condition* on intentional action, and I will assume it for the meanwhile, though remarks will be made in support of it in Section 4. If this genuinely ought to be a constraint that intentionalist theories of self-deception must respect, then it seems that philosophers who believe that intentional self-deception gets accomplished through the attentional strategy would not be worried by it.

Allow me to offer a conjecture as to why attentionalism is a particularly popular strategy theory among intentionalists. I suggest that this may be because *prima facie*, the idea of intentional self-deception being accomplished through the attentional strategy seems to escape the threat of paradox, unlike with some other prospective strategies of intentional self-deception, as I will explain.

What stands in the way of the subject achieving his goal of believing that p is the evidence warranting the belief that not-p (or his knowledge of it). So clearly something must be done about that. I suggest that ways the subject may deal with this threatening evidence can be placed into one of two categories: flight responses or fight responses. The attentional strategy is a flight response. With this, the self-deceiver turns away from the evidence, tries to ignore it, and tries to suppress his knowledge of it. A fight response, on the other hand, would involve confronting the evidence and trying to disarm its unwelcome import. With this sort of response, S remains aware of the evidence's existence, but that evidence is rendered innocuous. Note that these two sorts of response are mutually exclusive. The self-deceiver could no more indulge in both together than one could simultaneously flee from and fight a menacing animal.

I suggest that it is more difficult to see how one could successfully employ some of the traditional fight responses mentioned in the self-deception literature with the intention of deceiving oneself. Take, for instance, the typical fight response of attempting to explain away the evidence (also called 'rationalization'). Suppose that S adduces specious considerations C to undermine threatening evidence E.[7] Now it seems that S can put C forward in either the belief

---

[7] We must suppose that the considerations are specious. If they were valid considerations, S's adducing them would hardly count as an attempt to deceive himself.

that *C* genuinely throws *E* into significant doubt or in the belief that it is specious, and designed only to fool one into thinking that *p*. But if the former, why should we regard *S*'s adducing of *C* as an attempt at *deception*, and if the latter, though we could regard this as an attempt at deception, how would *S*'s opinion ever be swayed by it?

The attentional strategy does not seem so burdened with such problems. As we shall see, at the core of the attentionalist theory is the idea that by shifting our thoughts away from something, we can make ourselves forget about it. But the idea that we can shift our thoughts away from some matter with the intention of forgetting it and succeed is not paradoxical. It might seem paradoxical if we thought of the event of forgetting as an intentional action, but it is not; it is the intended result of an intentional action (shifting attention to something else). Analogously, the event of falling asleep is not an intentional action, though it may be the intended result of certain intentional actions (lying down, closing one's eyes, etc.). I suggest that the idea that we can make ourselves forget/not think about something is no more paradoxical than the idea that we can make ourselves fall asleep.

The examination of attentionalism would gain significance were it true that intentionalism's best prospects for avoiding the traditional charge of paradox lay with flight response theories like attentionalism (the only flight response strategy theory I am aware of). However, I offer this idea only as a conjecture for explaining attentionalism's popularity with intentionalists. Evaluating this conjecture would clearly take much further work and would require examining all other prospective self-deceptive strategies (fight responses in particular) and testing for paradox. Suffice to say that the fact that attentionalism is the most popular strategy theory among intentionalists is alone enough to give importance to this examination.

3. How is suppression supposed to result in belief change?

Let us return to considering how the attentional strategy is supposed to work. The assumption of the attentionalist account is that we can induce doxastic changes in ourselves by simply shifting our attention between belief-relevant considerations. But we may legitimately wonder about how this could ever be effective. The worry is well put by David Kipp:

. . . the main difficulty with attention-oriented theories is that they make plausible merely how a person might contrive to avoid thinking about an unwelcome belief . . . To put something one knows out of one's immediate consciousness by directing one's attention elsewhere can effect, so long as successful, only that one ceases to think about that thing, and not that one ceases to know it. (1980, 310)

So merely directing one's attention away from C^- would not, in Kipp's view, make one no longer know about that evidence and the truth it supports. It might relieve one of the anxieties associated with thinking about the unwelcome truth, for a while at least, but it would not relieve one of the belief in that truth itself.

However, many attentionalists seem to think that these shifts of attention can accomplish more than merely removing thoughts of the unwelcome evidence/belief from consciousness. Perring claims that '[i]ntentional self-deceivers manage to maintain the belief that not-*p* by intentionally ignoring, blocking out, and engaging in activities *that will lead them to forget* evidence that *p* . . . they find ways to forget that information' (1997, 124, emphasis added). Additionally, in Van Leeuwen's view, '[s]uppression may be used deliberately in willful self-deception to *undermine memory* of evidence contrary to the desired belief' (2008, 202, emphasis added). And Hamlyn, who also claims that intentional self-deception can be accomplished through the attentional strategy, says that, '[b]y putting something out of mind repeatedly a man may come as a result to forget that very thing' (1971, 46; also see Whisner 1998, 196). So according to these philosophers, averting attention from C^- is supposed to not just make one not think about C^-, but to forget about C^-.

(We might think that Davidson offers an alternative proposal in saying that the knowledge of the unwelcome evidence, or the belief this knowledge supports, gets *partitioned* or *walled off* as a result of the attentional strategy. However, Davidson's talk of partitioning is, as he admits ([1997] 2004, 220), *metaphorical*. Being metaphorical, it must be translated into more familiar terms that have a literal application, if fuzzy thinking is to be avoided on this matter. I am suggesting two ways in which this can be done. Either a belief being partitioned off means that the belief and the evidence for it are no longer on the subject's conscious mind ('not being accessed') or it means that this thought-content is forgotten ('inaccessible'). These are the

possibilities under consideration here and it is unclear how else Davidson's metaphor could be cashed out).

The idea that forgetting about the evidence which formed the basis for a belief could cause you to lose that belief sounds like one that we could grant the attentionalist. For instance, it has been observed (anecdotally) that peoples' low opinions of politicians or political parties can subside after time passes and they forget about the misdeeds which earned their contempt in the first place. Perhaps the more contentious point here is that these shifts of attention would *need* to result in this forgetting if the mentioned doxastic changes are to occur. I think that this is the case, but before I can explain why, it will be necessary to say a little about what it means to forget something.

When do we say that something has been forgotten? Consider Smith, who learned of a shortcut for getting from A to B. Now if the matter of the shortcut is not on his mind or in his thoughts, or even if he has not thought about it for years, we cannot conclude from that fact alone that he has forgotten about the shortcut. However, if it does not come to mind *at a time when that information is relevant*, then we have grounds for saying that he has forgotten it. For instance, if Smith is travelling from A to B then that is a time when that information is relevant, and if it did not enter his mind then (and he ended up taking the long route instead), we could affirm that he forgot all about the shortcut. Another such occasion would be if someone were to ask Smith, 'do you know of any shortcuts from A to B?' If the information did not come to mind here, we would also have reason to say he forgot it (note that though Smith might not remember the information, he might still remember *that he knows* it and might try, in vain, to recall it).

Consider now our self-deceiver. The self-deceiver has tried to make himself believe that *p*. Following Bach (1981) and McLaughlin (1988), I will assume that believing that *p* implies that when one consciously thinks of the *p*-issue, and of whether *p*, one will tend to consciously think that *p* is true.[8] And this is what he wants; if he thinks that not-*p* when he thinks of the *p*-

---

[8] Though it might be logically possible for someone who believes that p not to occurrently think that *p* when he thinks of the *p*-issue, there would have to be a special explanation for why this did not happen. People who believe that p are generally disposed to think that *p* when they think of this issue, though perhaps this disposition can be blocked somehow.

issue, he will suffer the distress it was his goal to avoid.[9] However, an occasion when he is thinking of the *p*-issue is an occasion where his knowledge of C^- *is relevant*. For C^- is evidentially relevant, and indeed, *ex hypothesi* evidentially crucial to this issue. Thus, if C^- does not come to mind on that occasion, it has been forgotten. But this is the precise occasion when C^- must not come to mind (if he is to think that *p*), given that he (*ex hypothesi*) took C^- to constitute compelling grounds for thinking that not-*p*. In short, he needs that information to stay out of mind when it is relevant to the occasion, which by our (hopefully reasonable) assumptions implies that he needs to forget that information.

But perhaps there is another way of understanding this. Perhaps the strategy works in the following way. By repeatedly shifting attention off C^- and onto C^+, one's estimate of the former's evidential weight decreases, while one's estimate of the latter's evidential weight increases. In that case, C^- would not need to be forgotten. It could still come to mind when the self-deceiver thinks of the *p*-issue, without this jeopardizing his thinking that *p*, since he now sees C^+ as carrying more weight. However, it is not clear how these attentional shifts could accomplish this. For merely attending more or less to evidence does not in itself seem to make one's *attitude* towards that evidence change. Consider, for instance, the Müller-Lyer illusion. The evidence of how the lines look may have originally led one to think that they are of unequal length, though afterwards, on the basis of having measured the lines say, one may believe that they are equal. But attending intently to the illusory look of the lines (evidence that *p*), at the expense of the memory of having measured them (evidence that not-*p*), would not seem to make this visual evidence seem more persuasive. (Though if one forgot about having measured the lines, one might revert to being persuaded by their illusory appearance.)

Rather, for one's attitude towards some evidence to change (for instance, from thinking that it warrants believing that not-*p* is true to thinking that it does not), one would have to reinterpret, re-assess, or re-evaluate it. For instance (and to borrow an example from Mele 1997, 94), if Roz refuses to date Sid, and Sid initially takes this as evidence that, contrary to his hopes, Roz does not fancy him, it is hard to see how averting his attention from the memory of this rejection will change his understanding of the significance of this fact. Rather, for that to change, a re-interpretation or re-evaluation of the incident would probably be necessary, say, by

---

[9] He might just want to avoid thinking of the *p*-issue entirely, but this would not be a case of trying to deceive himself in the mentioned sense.

supposing that it could be just a case of her playing hard to get. However, such actions (also called 'rationalizing' or 'explaining away' the evidence) require doing the opposite of turning attention away from the evidence. They require that we think about and therefore attend to it. Or in terms of the distinction made earlier, this would be a fight rather than a flight response, which are mutually exclusive responses. Therefore, this would not be a case of using the attentional strategy.

Let us now compose a more detailed description of how intentional self-deception is supposed to happen on the attentionalist account, incorporating all that has been said. Note that, as Whisner and Hamlyn suggest, on this view the self-deception may take time. It may involve repeated attempts at suppressing C^-, thus rendering C^- progressively less accessible to recall, finally resulting in the loss or 'burial' of the unwelcome belief.

> A person who encounters unwelcome considerations C^- which make her warrantedly believe that a distressing proposition not-p is true, but who wants to believe that p (to avoid the distress of believing that not-p) may achieve this by shifting her attention off C^-, and onto considerations C^+ which are supportive of the more agreeable proposition p. As a result (perhaps after repeated attempts), she ends up forgetting C^-, and consequently she loses the (conscious[10]) belief which C^- supports. Because now C^+ is all she remembers, the considerations she has to judge from weigh in favour of p, and she acquires that belief.

There seems to be nothing straightforwardly paradoxical or incoherent in the idea that this could work. There are, however, reasons to deny that this is a *psychologically* possible phenomenon, relating to its assumptions concerning the effectiveness of repressive strategies. In the following section I will outline these objections.

---

[10] I put '(conscious)' here since instead of speaking of S losing the belief, one might want to speak of S as rendering that belief 'unconscious'/'inaccessible'. These different ways of speaking may be intended to reflect the difference between knowledge/information being erased from memory and being just forgotten without being completely erased. So here we would be trying to emphasize that it would not be impossible for S to eventually recall C^- and regain the conscious belief that not-p.

4. Can suppression result in repression?

Kipp anticipated the response that these acts of shifting attention are supposed to undermine the unwelcome belief by making us forget what we know, though he found the idea psychologically unrealistic (also see Martin 1998, 191–192). But this view is drowned out in the philosophical literature by more frequent assertions that such a thing can happen. At this point, philosophical debate reaches its limit and the question of whether self-deception, as the attentionalist sees it, is possible comes within the purview of psychology. For it is an empirical question whether intentionally shifting your attention off C^-, repeatedly perhaps, can lead to you forgetting C^-. Unfortunately, most attentionalists have overlooked the relevant empirical work on this issue. To this we will now turn.

There have by now been many studies done investigating thought-suppression, and some of these have studied the effects of suppression on accessibility of the suppressed item to later memory recall. Some experiments tested the effect of suppressing affectively neutral memories on their accessibility for recall and found modest reductions in their accessibility (Anderson and Green 2001)[11]. However, later attempts to replicate these results failed, finding suppression to have no negative effect on later recall ability (Bulevich et al. 2006).[12] Of more relevance to the self-deception debate, however, are studies that focused on distressing thoughts. Rassin, Merckelbach, and Muris (1997) conducted an experiment where subjects watched a disturbing film fragment of a man being attacked by a bear. They were sent home, but had to return to the laboratory five hours later, with half of them instructed to suppress thoughts of the clip in the

---

[11] Van Leeuwen, the only philosopher to have referred to the empirical work on thought-suppression, claims that self-deceivers employ repressive strategies to achieve their end, and cites this experiment as 'put[ting] to rest' (2008, 202) the issue of whether repression is possible. It's regrettable that he just examined this article, since it's unrepresentative of the field of study as a whole.

[12] These experiments consisted of three stages. In the first stage, subjects had to learn unrelated word-pairings (e.g. ordeal-roach). In the second stage, they were presented with the first 'cue' words on a screen, and had to either call out the associated word, or in other cases suppress the thought of it. In the third stage, to test the effect of suppression on recall, subjects were given the cue words (ordeal-?) and had to recall the associate, whether previously instructed to suppress it or not. Or they would be given an associated word with the first letter of the target word (insect-r_) and had to complete the word.

meanwhile. On their return, they completed a memory test on the details of the clip. Suppressers' memory was found to be no worse than that of non-suppressers.[13]

In fact, not only has little evidence been found that suppression can lead to forgetting to any significant degree, but research has highlighted the difficulty involved in simply keeping our minds off intrusive or unwanted thoughts. As Wegner and co-workers remark, the thought-suppression literature 'indicates that people have trouble even eliminating a thought from consciousness, much less erasing it from memory' (Wegner, Quillian, and Houston 1996, 680). The most remarkable finding here is probably what is termed the 'ironic' or 'paradoxical effect' of thought-suppression: the so-called 'rebound effect'. This refers to the fact that attempting to avoid thoughts tends to make those thoughts intrude on one's mind even more.

For instance, in one experiment which tested the long-term effect of suppression on thought-frequency over a seven-day period, subjects were instructed to think for two minutes about the most distressing event they had experienced during the last two years, and then to keep this memory suppressed as best they could for two minutes. It was found that the subjects most successful at rebuffing these thoughts during this time were harassed the most with intrusions from them over the following seven days, thus demonstrating the 'rebound effect' over a long-term period (Geraerts et al. 2006). In another study, subjects identified a negative intrusive thought they had experienced and dwelled on it for a few minutes. They were then sent home and asked to record for four days whenever the thought came to mind, though some were instructed to do their best to keep it suppressed from consciousness over that period and others told to just take note of it. But despite (and because of) their efforts, suppressers recorded over twice the amount of intrusions from those thoughts as non-suppressers (Trinder and Salkovskis 1994). The rebound effect has been found to be a quite robust phenomenon (see Rassin 2005). '[I]t is safe to conclude' writes Rassin in a literature review, 'that suppression is ineffective in the short run, and even has a paradoxical rebound effect in the longer run' (Rassin 2005, 54). So it is not just that attempts at suppressing thoughts do not lead to our forgetting them. In Wegner and Erber's

---

[13] Suppressers also had more thoughts about the clip over the five-hour period, manifesting suppressions 'ironic effect', which will now be discussed.

words, such attempts often have the opposite effect, rendering them 'hyperaccessible' to consciousness (1992).[14]

The question now is whether we can conclude from these studies that in the attentionalist cases, the rebound effects of suppression would intervene and scupper the attempt to deceive oneself. The attentionalist may deny that we have good reasons for concluding this by arguing that there are crucial differences between what is happening in these experimental cases and the real-life cases he has in mind, which prevent us from legitimately extrapolating from the one to the other.

First, it could be objected that these experimental situations are so artificial and unnatural that we cannot legitimately draw conclusions from them about whether suppression is effective 'outside the laboratory'. For in these experiments, subjects are under explicit instruction to take note of occurrences of the thought they must suppress, unlike in real-life cases (by, for example, pressing a button in short-term studies or noting them in a diary in long-term studies). But their awareness of being in this experimental situation, and under such instructions, may serve as a constant reminder of the thought they are supposed to not think about, making suppression artificially more difficult. However, we should bear in mind that in these experiments, suppressers experienced more occurrences of the relevant thought than people in the control groups, who were also under instructions to take note of occurrences of that thought, though they were not instructed to suppress it. So even if suppression is made artificially more difficult in these experimental conditions, we can still conclude that it does not work.

Second, the attentionalist may insist that the thoughts involved in the real-life cases he has in mind would be more anxiety provoking than the ones in these studies, and consequently, that subjects will be more motivated to succeed in suppressing them. But there is no reason to think that this should make any difference to how successful one would be at thought-suppression. For a start, it is well known that greater affective significance associated with a memory or thought is generally proportional with its greater vividness, longevity, and accessibility to consciousness (e.g., Walker, Rodney, and Thompson 1997). And regarding motivation to succeed, the lesson of the rebound effect seems to be that the more one tries to

---

[14] The counterproductive nature of attempts to suppress unwanted thoughts has been known to some psychiatrists for some time, such as those working with war veterans with traumatic memories. See Rivers (1920).

suppress thoughts, the less successful one tends to be, so more motivated subjects should not do any better.

Third, it could be objected that in these experimental situations, the subjects were explicitly instructed to suppress certain thoughts and proceeded to carry out this instruction in a conscious, deliberate manner, while in the attentionalist cases it is more probable that these actions are done unconsciously, which is why they would be more successful. However, this response is not available if it is true that, as was claimed earlier, doing something intentionally implies doing it knowingly. This is a large issue and I will not be able to fully defend this thesis here, but allow me to make some remarks in support of it.

Many philosophers have held that knowing that your action falls under a certain description is essential for the action to be intentional under that description (e.g. Anscombe 1966, 11, 87). The plausibility of this may be brought out from the following. Consider these different descriptions of an action that I may perform at an auction: (1) I raise my finger, (2) I bid for the painting, and (3) I bid for the forgery. Let us say, however, that I did not know that the painting was a forgery, though I knew that I was raising my finger and bidding for a painting. In that case, we would judge that I did not, and indeed could not have intentionally bid for the forgery, though I knew that I was bidding for a painting and bid for it intentionally, and the explanation for this judgement seems to be that knowledge of what you are doing is necessary if you are to do it intentionally. Furthermore, if the studies of Malle and Knobe (1997, 112–114) are anything to go by, this judgement would be in accord with those of 'lay people' too. Malle and Knobe found that when certain scenarios were described to 225 undergraduate students involving people performing certain actions, they were generally unwilling to attribute intentionality to the actions if the agents were described as being unaware that they were doing them. They conclude that to the theorist 'who believes in unconsciously performed intentional actions . . . the results must be surprising, if not damaging' (1997, 114).

The three most important proposals for counter examples to this knowledge condition that I have come across are as follows. First, consider a case where the success of our intentional enterprise is not a sure thing (Davidson [1978] 2001, 91–92; Ross 1982, 264). Imagine NASA launching a rocket programmed to destroy a huge meteoroid hurtling towards Earth. Imagine also that they have serious doubts that the mission will succeed though the exigency demands that they try, and suppose further that they do succeed. Because of these doubts, the argument

goes, we may not be able to describe them as having known their action would result in the destruction of the meteoroid (if such uncertainty is incompatible with knowledge), yet it seems correct to say that they intentionally destroyed the meteoroid. This may be an important exception to the knowledge condition, yet we would be overreacting to it if we were to deny, on its basis, any intimate link between intentionally action and awareness of what you are doing, since surely in all cases like this, the agents would have to be aware that they were at least trying to do what they succeeded in doing. It is hard to imagine how NASA, for instance, could have intentionally destroyed the meteoroid without having known that that is what they were trying to do. Exceptions like this are limited to cases where there is doubt as to whether an action will succeed, and they do not prove the possibility of unconsciously intentional action in the sense congenial to the attentionalist here.

Second, Ross considers a purposive action done 'by rote' and 'mechanically', like shifting into third gear in the course of driving, and says, '[m]y intuition is that such an act is intentional even though one does not at the time [know] . . . that one has performed it' (1982, 263) and others too have expressed similar views (e.g., Martin 1997; Fingarette 1998). I presume that the reason why Ross assumes that in such cases the subject did not know at the time that he was doing this action is because he is imagining that the subject was not aware of himself doing the action, his attention being engrossed in other things. However, whether people really are unqualifiedly unaware of what they are doing in such cases is debatable. For awareness, it seems, need not be an all or nothing affair, but may come in degrees, and it seems that such cases would not untypically involve partial or peripheral awareness of what one is doing (that is, it seems as wrong to say, unreservedly, that subjects are completely unaware and oblivious to what they are doing in such cases as it is to say they are fully aware of what they are doing). Granted, in light of such peripheral awareness we would be as reluctant to say, unqualifiedly, that the subject knew she was shifting into third gear as we would be to say that she did not know this, but this reluctance is proportional with, and explains, our hesitation on the question of whether to classify this as an intentional or unintentional action. It is an exemplar of neither. This case does not clearly demonstrate a disconnection between intentional action and knowledge of what you are doing.

Third, Bermúdez proposes that cases where we 'lose touch' with an intention while in the process of implementing it, when the implementation process is long and drawn out, illustrate

that we can do things intentionally yet unknowingly (2000, 314). An example would be that of a wage earner who works with the intention of earning a wage, but while going about his work, does not remain conscious of this intention. He therefore, in Bermúdez's view, would be unknowingly acting with the intention of earning a wage (314). However, Bermúdez implicitly foists on his opponent an unnecessarily strong claim here: that an agent must remain conscious of an intention over a period to be knowingly acting with that intention over that period. But knowing something does not entail being conscious of it. The wage earner still knows his intention in doing what he is doing simply because if you were to ask him what his intention is in doing it, he could tell you straight off that it is to earn a wage. To know his intention in what he is doing, he need not be constantly thinking of it; he just should not have forgotten it and should be able to tell us what it is if asked.

I will not pretend to have satisfactorily defended the knowledge condition thesis in these few pages, but if it ultimately turns out to be true, then the fact that the subjects in the thought-suppression experiments were consciously trying to suppress certain thoughts should not differentiate them from the subjects in the attentionalist cases. Furthermore, it is not clear that there are any other important differences between what subjects are doing in these experiments and what subjects are supposed to be doing in the attentionalist cases, which would bar us from extrapolating from the former to the latter.

5. Conclusion

To sum up, attentionalists should be aware that there is plenty of evidence suggesting that we do not have the kind of mental control over ourselves which attentional accounts of self-deception are premised on. Notwithstanding the would-be self-deceiver's efforts to ignore her knowledge of the evidence for not-*p*, that knowledge will remain accessible to her consciousness, so that as soon as she considers the *p*-issue, that knowledge will be there to influence how she thinks about the matter. All in all, there is very little reason to think that doxastic changes can be effected by a person's attempts to shift attention from thoughts of evidence. There is little reason to think that intentional self-deception, on the attentionalist account, is psychologically possible.

Acknowledgements

Notes on contributor

Kevin Lynch is an Irish Research Council Postdoctoral Fellow at University College Dublin. He has published a number of articles mainly relating to self-deception.

References

Anderson, M.C., & Green, C. (2001) 'Suppressing Unwanted Memories by Executive Control', *Nature*, 410(6860), 366-369.

Anscombe, G.E.M. (1966) *Intention*, Oxford: Basil Blackwell.

Audi, R. (1976) 'Epistemic Disavowals and Self-Deception', *The Personalist*, 57, 378-385.

Bach, K. (1981) 'An Analysis of Self-Deception', *Philosophy and Phenomenological Research*, 41, 351–370.

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

Bermúdez, J. L. (2000) 'Self-Deception, Intentions, and Contradictory Beliefs'. *Analysis*, 60: 309–319.

Bulevich, J.B., Roediger III, H.L., Balota D.A., & Butler A.C. (2006) 'Failures to Find Suppression of Episodic Memories in the Think/No-Think Paradigm', *Memory and Cognition*, 34, 1569-1577.

Davidson, D. (2004/1986) 'Deception and Division' in *Problems of Rationality*, Oxford: Clarendon Press, pp.199-212.

Davidson, D. (2004/1997) 'Who is Fooled?' In *Problems of Rationality*, Oxford: Clarendon Press, pp.213-230.

Davidson, D. (2001/1978) 'Intending', in *Essays on Actions and Events*, Oxford: Oxford University Press, pp.82-102.

Fingarette, H. (1998) 'Self-Deception Needs No Explaining', *Philosophical Quarterly*, 48, 289-301.

Geraerts, E., Merckelbach, H., Jelicic, M., & Smeets, E. (2006) 'Long-Term Consequences of Suppression of Intrusive Anxious Thoughts and Repressive Coping', *Behaviour Research and Therapy*, 44, 1451-1460.

Hamlyn, D.W. (1971) 'Self-Deception', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 45, 45-60.

Kipp, D. (1980) 'On Self-Deception', *Philosophical Quarterly*, 30, 305-317.

Lockie, R. (2003) 'Depth-Psychology and Self-Deception', *Philosophical Psychology*, 16, 127-148.

Malle B.F., & Knobe, J. (1997) 'The Folk Concept of Intentionality', *Journal of Experimental Social Psychology*, 33, 101-121.

Martin, M.W. (1997) 'Self-Deceiving Intentions', *Behavioral and Brain Sciences*, 20, 122-123.

Martin, T. (1998) 'Self-Deception and Intentional Forgetting: A Reply to Whisner', *Philosophia,* 26, 181-194.

McLaughlin, B. (1988) .Exploring the Possibility of Self-Deception in Belief'. In Perspectives on Self-Deception, edited by B.P. McLaughlin and A.O. Rorty, Berkeley: University of California Press (pp.29–62).

Mele, A. (1997) 'Real Self-Deception', *Behavioral and Brain Sciences*, 20, 1997, 91-102.

Perring, C. (1997) 'Direct, Fully Intentional Self-Deception is Also Real', *Behavioral and Brain Sciences*, 20, 123-124.

Pugmire, D. (1969) '"Strong" Self-Deception', *Inquiry*, 12, 339-346.

Rassin, E. (2005) *Thought Suppression*, Amsterdam; London: Elsevier.

Rassin, E., Merckelbach, H., and Muris, P. (1997) 'Effects of Thought Suppression on Episodic Memory', *Behaviour Research and Therapy*, 35, 1035-1038.

Rivers, W.H.R. (1920) *Instinct and the Unconscious*, Cambridge: Cambridge University Press.

Ross, G. (1982) 'Knowledge and Intentional Action', *Philosophical Studies*, 41, 263-266.

Trinder, H., & Salkovskis, P.M. (1994) 'Personally Relevant Intrusions Outside the Laboratory: Long-Term Suppression Increases Instrusion', *Behaviour Research and Therapy*, 32, 833-842.

Van Leeuwen, D.S.N. (2008) 'Finite Rational Self-Deceivers', *Philosophical Studies*, 139, 191-208.

Walker, W.R., Rodney, J.V., & Thompson, C.P. (1997) 'Autobiographical Memory: Unpleasantness Fades Faster Than Pleasantness Over Time', *Applied Cognitive Psychology*, 11, 399-413.

Wegner, D.M. (1994/1989) *White Bears and Other Unwanted Thoughts*, New York; London: The Guilford Press.

Wegner, D.M., Quillian, F., & Houston, C. (1996) 'Memories Out of Order: Thought Suppression and the Disturbance of Sequence Memory', *Journal of Personality and Social Psychology*, 71, 680-691.

Wegner, D.M. & Erber, R. (1992) 'The Hyperaccessibility of Suppressed Thoughts', *Journal of Personality and Social Psychology*, 63, 903-912.

Whisner, W. (1998) 'A Further Explanation and Defense of the New Model of Self-Deception: A Reply to Martin', *Philosophia*, 26, 195-206.