

Self-deception and shifting degrees of belief

Chi Yin Chan and Darrell P. Rowbottom

darrellrowbottom@ln.edu.hk

A major problem posed by cases of self-deception concerns the inconsistent behavior of the self-deceived subject (SDS). How can this be accounted for, in terms of propositional attitudes and other mental states? In this paper, we argue that key problems with two recent putative solutions, due to Mele and Archer, are avoided by “the shifting view” that has been advanced elsewhere in order to explain cases where professed beliefs conflict with actions. We show that self-deceived agents may possess highly unstable degrees of belief concerning the matters about which they are self-deceived.

1. Introduction

Consider the following scenario presented by Archer:

[A] man ... 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. (2013, p. 265)

Situations like this are reasonably common. An alcoholic might claim that his drinking is under control, but refuse to discuss his alcohol intake with doctors. And a cancer victim might deny that he suffers from cancer, but ask his distant friends and family to visit him soon. The key feature shared by these cases is that they involve a conflict between the subject's verbal behavior and nonverbal action.¹ So it is hard to say whether the subject genuinely believes in the professed belief or not. Furthermore, as Porcher (2012) emphasizes, the nonverbal behavior will also be inconsistent in some cases. That is to say, the SDS's nonverbal behavior will sometimes suggest that they possess the undesirable yet warranted belief, and sometimes suggest that they possess the desirable yet unwarranted contrary belief. Consider the case of the cancer victim again. In addition to asking his distant friends to visit him as soon as possible, he might make a detailed plan about his life after retirement and join a package tour starting next summer (although he has been told that he cannot live that long).

Several different theories for explaining self-deception have been proposed. Intentionalists, such as Davidson (1985), hold that we should model the phenomenon on interpersonal deception. Such cases involve a deceiver who believes that $\sim p$ and who convinces a victim, the deceived, that p . This raises two difficulties, however, which are known as the static and dynamic paradoxes. The static paradox occurs because in order to simultaneously play the role of the deceiver and the role of the deceived, the SDS must simultaneously hold two contrary beliefs (and be aware of so doing). But that appears, at least *prima facie*, to be an impossible state of affairs. The dynamic paradox occurs because in order to deceive and be

¹ Not all cases involving conflicts between verbal behavior and nonverbal action are cases of self-deception, however. Schwitzgebel (2001; 2010) presents several cases where self-deception is clearly absent, such as the Antonio case discussed later in the paper.

deceived, the SDS must be aware and unaware that attempted deception is underway. And that also appears to be an impossible state of affairs.

Intentionalists typically attempt to avoid these paradoxes by “partitioning”; they argue that the attempt to deceive may temporally precede the deception, or that different psychological parts of an SDS serve as deceiver and deceived. The temporal route is plausible in some cases, such as when a person forgets that they intended to deceive themselves by previous actions and changes their belief as a result, as Sorensen (1985) points out. As Levy (2004) argues, however, these are cases of a special kind. Indeed the example with which we began, concerning the self-deceiving husband, is not of such a kind; there is a “back and forth” or vacillating aspect to the behavior involved. The psychological route is taken, in different ways, by Davidson (1982), Pears (1984), and Rorty (1988). It has better potential for generalizing intentionalism, although it has been criticized by the likes of Mele (2001; 2012), Galeotti (2012), and Scott-Kakures (2012). We won’t explore it herein, as the current debate tends to concern non-intentionalist, or revisionist, alternatives.²

There are two revisionist approaches. The first “deflationary” option involves attribution of only one belief of the contrary pair to the SDS. The second “non-doxastic” option avoids the attribution of either of the contrary beliefs. In the next section, we will consider one representative version of each approach in greater detail. We will argue that each encounters problems which our preferred view, “the shifting view”, avoids. In the section thereafter, we will provide an overview of the shifting view and address some misconceptions concerning it.

² However, it is possible that an account that appeals to shifting of degrees of belief is compatible with intentionalism. For instance, a temporal partitioning approach could posit multiple small shifts in degrees of belief.

In the penultimate section, we will show how the shifting view can explain the behavior of the SDS in something like the hypothetical case with which we began. We will conclude with a brief summary of our findings.

Before we continue, a concise sketch of the shifting view will be useful. On this view, there are situations in which individuals' degrees of belief are highly sensitive to relatively subtle changes in context, although the "shifting" patterns involved may be relatively stable (and hence predictable). Thus a person might profess belief in a proposition in one class of contexts because she believes it in those contexts, and behave as if she doesn't believe it in another class of contexts because she doesn't believe it in those contexts. The changes might also be less dramatic. High confidence might be apparent in assertion, and middling confidence might be apparent in action, for instance. This is not to say that all cases where professed beliefs conflict with actions are explicable by appeal to the shifting view (on a factive view of explanation, at least).

2. Revisionist Avenues

2.1 Mele's Deflationary Approach

Mele's (2001; 2012) deflationary account of self-deception is one of the most well-known and widely discussed in contemporary philosophy. The main idea behind it is that self-deception is a purely causal mechanism in which the SDS is driven by a desire, or desires, to "acquire the belief in a suitably biased way" (Mele, 2012, p. 2). Initially, Mele added to this that the "sufficient conditions for entering self-deception" in acquiring the belief that *p* are as follows:

1. The belief that p which S acquires is false,
2. S treats data relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way,
3. This biased treatment is a nondeviant cause of S 's acquiring the belief that p , and
4. The body of data possessed by S at the time provides greater warrant for $\sim p$ than for p .

(Mele, 2012, pp. 2–3)

But consider again the case of the husband whose wife is unfaithful. His unusual behavior requires him to recognize the ugly fact, or at least suspect that his wife is being unfaithful, otherwise he would not perform the avoidance behavior which is inconsistent with his professed belief. As it stands, however, Mele's view doesn't require the SDS to have any doxastic (or relevant sub-doxastic) attitude towards the unwanted proposition that $\sim p$. Hence, it fails to account for the inconsistency of self-deception.³

Mele (2012, p. 12) therefore added that “ S consciously believes at the time that there is a significant chance that $\sim p$ ” should be included as one of the (jointly) sufficient conditions.

³ Note also that the fourth condition doesn't require that S must be aware that the evidence available provides greater support for $\sim p$ than for p . And if the SDS doesn't recognize that the evidence available provides greater support to the undesirable proposition that $\sim p$, she has no reason to “deceive” herself.

This suggests the SDS believes that p , although she also strongly suspects the converse.⁴ But as Porcher asks:

If a person believes that p but at the same time isn't quite sure or "suspects" otherwise (i.e., believes that there is evidence that not- p), should we attribute to her a pair of contradictory beliefs (albeit with different degrees of confidence) or just one belief that p with a degree of confidence below 1? (2012, p. 76)⁵

The problem with the first option is that it results in a new variant of the static paradox. Yet, as we saw previously, the main motivation for adopting a revisionist approach to self-deception is to avoid such a paradox. Hence, it's extremely unlikely that Mele would want to take this route if it can be avoided. The second route is much more attractive.

It should be noted before we continue, however, that there is a third option which Porcher doesn't consider. Mele's view is also compatible with maintaining that the SDS (fully) believes p and (fully) believes "The probability of not- p is greater than n ," where n is an appropriate threshold. That's because having a belief that the probability of p has a particular value is not equivalent to (and does not entail) having a degree of confidence in p which has that value. One reason this is true is that many different interpretations of probability are possible, and interpreting probabilities as (rational) degrees of belief is just one option among

⁴ As Porcher (2012) points out, if Mele were not to understand "doubt" or "suspicion" in doxastic terms, he would have to explain how the subdoxastic attitude involved could "override" the desirable belief and generate the SDS's behavior.

⁵ Belief that "there is evidence that not- p " is the locution of Mele (2001, p. 76).

many.⁶ Interpreting them instead as being “in the world”—as being relative frequencies in the limit or propensities—is also a live option.⁷ Moreover, this is suggested by Mele’s (2012, p. 12) use of ‘chance’ rather than ‘probability’.

Many formal epistemologists think that a subject’s credences (or degrees of confidence) should match their estimates of chances (*qua* propensities or relative frequencies in the limit). But it is straightforward to comprehend how they might fail to do so. For example, an obsessive gambler who is well-versed in probability theory might see the ball land in a black compartment on twenty repeated spins of a roulette wheel, yet nonetheless be confident—say, on the basis of (what she’d describe as) “a gut feeling”—that it will land in a red compartment on the next spin. Irrational she might be, but no static or dynamic paradox arises.

Nevertheless, there is an excellent reason to avoid the third route. In order to have a belief with the content “The probability of not- p is greater than n ” (or some appropriate surrogate), a subject must grasp the concept of probability. (A subject would also have to grasp the more specific concept of world-based or aleatory probability in order to have a belief concerning a chance, and so forth.) However, it is reasonable to think that self-deception occurred before probability, in anything akin to the modern sense, had been conceived of. It is also plausible that children (and others in the modern world) who haven’t grasped the concept of probability are capable of self-deception. This would evidently be true if self-deception has “evolved to

⁶ Moreover, having a belief that one’s degree of belief in p is greater than n doesn’t entail that one’s degree of belief is greater than n . There are good reasons to think that degrees of belief aren’t transparent, as explained in Rowbottom (2016). We touch on these later.

⁷ See, for instance, Gillies (2000) and Rowbottom (2015).

facilitate interpersonal deception by allowing people to avoid the cues to conscious deception that might reveal deceptive intent,” as von Hippel and Trivers (2011) contend.

Thus, only the second option—holding that the SDS has a belief in p with a degree of confidence lower than 1—is viable as a means of precisifying Mele’s proposal.⁸ Hence, Mele should accept that degrees of confidence are relevant in explaining the phenomenon of self-deception. Once this is admitted, however, it is also natural to accept that changes in the SDS’s degree of confidence in p occur, in order explain some of the SDS’s different behavior in different contexts. For instance, it might be admitted that one believes in p provided one’s degree of belief is greater than some threshold value, such as 0.5, and suspects that not- p provided one’s degree of belief is lower than another threshold value, such as 0.95.⁹ Then one might allow that an SDS concerning p may exhibit fluctuations of degree of belief in p between those two limits. When doubting “I have terminal cancer” more strongly, perhaps as a result of a tarot reading, a cancer victim might make plans for a holiday further into the future than she has repeatedly been told, by doctors, that she will live. But after a palliative

⁸ This entails having a degree of confidence in $\sim p$ of higher than 0, on the assumption that agents’ synchronic (active) degrees of belief in p and $\sim p$ sum to unity. If this condition doesn’t hold, something akin to the static paradox arises: it seems incoherent to think that one could, simultaneously, be confident in something being true and confident in the very same thing being false. This is not to deny that there are cases where agents take p and q to be possible (in conjunction) although p entails $\sim q$. These can be explained in virtue of agents failing to recognize that p and q are inconsistent, in virtue of p and q not being identical.

⁹ The thresholds may be context sensitive, if a connection between belief and degrees of belief is to be maintained, for reasons explored by Leitgeb (2014). For present purposes, however, it suffices to proceed without introducing this complication in the main text.

chemotherapy session, she might spontaneously burst into tears at the thought of her children growing up without her.

This is already to adopt a shifting view, albeit one where the extent of shifting—but not its frequency—is relatively limited. So the question now arises as to why it is worth maintaining that the shifting involved in self-deception must be limited such that the SDS always believes that p . Why not allow instead for the lower limit of degree of belief to be such that the SDS believes not- p some of the time? Neither the static nor the dynamic paradoxes present obstacles to thinking so. Furthermore, allowing for shifts to believing that not- p has explanatory benefits. Imagine now a hypochondriac, who has recently diagnosed herself as suffering from terminal cancer although she is in excellent health and has hastily written her will as a result.¹⁰ Upon being told by a “doctor of integrative medicine”—a quack—that her “cancer” is easily curable with the correct dietary changes, she might tear up the will in relief. It is hard to square such behavior with a mere suspicion on her part that she doesn’t have terminal cancer, however strong said suspicion might be. *Prima facie*, at least, it seems like the suspicion becomes rather more (although it will in all likelihood become a mere suspicion again at a later point). And since suspicion is construed as a degree of belief in not- p , it’s natural to think this has gone over a threshold such that belief in not- p is temporarily present.

Interestingly, however, Porcher (2012) argues that Mele’s view of self-deception is subject to a *reductio ad absurdum* in so far as it implicitly requires shifting (of broad scope) to occur.

He writes:

¹⁰ We use this example to indicate how the shifting view can handle cases of so-called “twisted self-deception” (Mele, 1999) where the SDS does not desire that p (but instead desires that $\sim p$).

Supposedly, when the subject manifests *p*-behavior, we would attribute to her the belief that *p*. When she manifests not-*p*-behavior, we would attribute to her the belief that not-*p*. This doesn't ... imply that the subject holds the belief that *p* and the belief that not-*p* simultaneously. However, this is ... a complete breakdown of the ordinary way of understanding and practicing belief attribution. Such an attribution has absolutely no explanatory or predictive power and makes out "belief" to be purposeless. [*italics added*] (Porcher, 2012, pp. 77–78)

We will return to this criticism of the shifting view—which originates in Schwitzgebel (2010) and also appears in Borgoni (2016) —in Section 3. But let us answer it briefly before we continue, by pointing out that the explanatory and predictive power of belief ascriptions is always limited in several respects. First, the predictive or explanatory power of the ascription of belief is hardly ever independent of the consideration of other psychological and situational factors, such as other propositional attitudes. That, indeed, is why behaviorism was a failed project and why dispositionalism about belief can't involve accounting for dispositions to act *solely* in terms of beliefs. For example, it doesn't follow from the fact that a person possesses a belief that "Line one is the same length as line two" that she will answer the question "Is line one the same length as line two?" in the affirmative. The Asch (1951) conformity experiments show precisely that other psychological factors, such as the desire to conform or the fear of unwanted attention, are relevant in predicting action. But this doesn't make the attribution of belief to the person in these experiments lack any explanatory or predictive power. On the contrary, we can predict how the person will behave in other circumstances, such as when asked what they think about the relative lengths of lines one and

two in a context in which they desire to tell the truth and there is no obstacle to them so doing.

Second, the predictive and explanatory power of ascribing beliefs is also limited in so far as beliefs can change. But as before, this doesn't make ascribing them "purposeless". A cheating husband is painfully aware that his wife will no longer believe he is faithful if she spots him kissing his lover, or if his lover calls his wife to detail his infidelities, for example.

Third, appeal to belief (and degrees thereof) in cases of self-deception has some predictive and explanatory power provided the beliefs in question (and the degrees thereof) aren't so sensitive to change that they alter whenever *almost anything else* changes. Think again of the hypochondriac as she's tearing up her will. At that point in time, she might also be liable to take steps to make the dietary changes specified by the quack. She might refuse to have a takeaway pizza if unexpectedly invited to do so by a friend, empty her fridge of particular foods if she happened to spot they were present, and so on. The fact that she's an SDS means only that some of her beliefs (and degrees thereof) are *considerably less stable* than we'd otherwise expect—that the aforementioned beliefs (and degrees thereof) are likely to change on the basis of triggers that would not alter the beliefs (or the degrees thereof) of a subject that wasn't in a self-deceived state (and was otherwise functioning in a mentally unimpaired, or statistically normal, way).

In summary, the attribution of belief (or of a particular degree of belief) in cases of self-deception does have predictive and explanatory power. The predictive and explanatory power is only diminished by—not eliminated in virtue of—the presence of self-deception. Self-deception only makes the beliefs or degrees of belief more liable to change.

2.2. Archer's Non-Doxastic Approach

Unlike advocates of deflationary accounts, such as Mele, Archer (2013) holds that we can explain the strange behavior of an SDS—at least in “the paradigm case” (Archer 2013, p. 281) —without holding that an SDS believes p or its contrary. Her strategy is to appeal instead to psychological categories other than belief.¹¹ More particularly, Archer's (2013, p. 276) proposal is that: “a *suspicion* that p , combined with *anxiety* that p , and a *desire* to believe $\sim p$ may meet the explanatory charge in some [paradigm] cases.”

To be more specific, Archer (2013) argues that a (paradigmatic) SDS is motivated by her desire and hope to engage in biased evidence gathering which stops her from holding the undesired belief that $\sim p$ while her “niggling doubts and suspicions” also prevent her from attaining the desired belief that p (Archer, 2013, p. 279).¹² As a result, Archer concludes that the attribution of belief in p or $\sim p$ to the SDS should be abandoned completely in accounting for self-deception. Her account of the example with which we began is as follows:

[T]he 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

¹¹ Non-doxasticists may instead claim that the SDS merely holds certain belief-like attitudes such as “alief”—a type of mental state introduced by Tamar Gendler (2008a; 2008b)—which are responsible for the inconsistent behavior.

¹² Biased evidence gathering includes actions such as avoiding information that favors the feared proposition and searching for evidence against the unwanted truth.

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. (Archer, 2013, p. 279)

Archer's non-doxastic approach is attractive in several respects. It is straightforward, backed by folk psychology, and unsusceptible to the static and dynamic paradoxes. *Prima facie*, it also appears to avoid any appeal to beliefs—other than background beliefs—whatsoever. However, Archer notes—following Lynch (2012; 2016)—that the attitude of suspicion might involve degrees of belief in some sense. She recognizes that “one may insist that a suspicion that *p* is or entails a low degree of belief that *p*” (Archer, 2013, p. 272).¹³ Archer attempts to resist this move, but grants “we ought to accept that a suspicion that *p* involves (although it

¹³ As intimated in our earlier discussion of suspicion in Mele's account, we hold that to suspect that *p* is to have a degree of belief that *p* within a particular interval. An interval seems necessary because, as Lynch (2016, p. 509) notes: “suspicion that *p* ... leaves room for a range of cases, since suspicion comes in degrees”. It is natural to think the upper limit of a suspicion that *p* is a degree of belief just below the threshold for counting as (or prompting) a belief in *p* (in the relevant context). The lower limit is harder to specify, but should evidently always be somewhat greater than zero. One don't suspect one's spouse of having an affair if one is extremely confident that—or if one is certain that—said spouse is not having an affair.

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 ’.” (Archer, 2013, p. 273). Archer’s idea is that no degree of belief that p (or corresponding degree of belief that $\sim p$) need be attributed to an SDS, although an SDS should be understood to hold one or more complete beliefs with p in their content. She writes:

[T]he self-deceived person often ... perhaps unconsciously ... 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 ... there is no explanatory need to attribute them the belief that ‘ p ’ [or that not- p], in order to explain their behaviour. (Archer, 2013, p. 273)

Before we criticize Archer’s account, we should remark on the use of ‘often’ in this passage. This might be present because Archer’s account only covers what she calls paradigm cases of self-deception, such as the one with which we began the paper. Or Archer might instead think there are some paradigm cases of self-deception where no such beliefs are present. She does not present or mention any such putative cases, though, so we will not pursue this potential avenue in any depth. We will say only that it seems like a difficult avenue to take because the presence of any of several different kinds of propositional attitudes concerning p appears to entail the presence of a degree of belief in p or beliefs with p in their content. Indeed, this seems to hold for the other propositional attitudes mentioned by Archer. For example, having a hope that p or an anxiety that p appears to entail having a non-zero degree of belief that p is true or a belief (or reasonably high degree of belief) that p is epistemically possible. One does not hope to win the first and second prizes in a lottery when one is certain that one only has one lottery ticket and that each ticket can only win one prize.

This brings us to the problems with Archer's account. First, it is less economical than positing a pair of degrees of belief (in p and $\sim p$) in any given time slice, in so far as it involves positing multiple beliefs involving p as well as several other propositional attitudes involving p .¹⁴ And it is typical to prefer the simpler of two potential explanations, *ceteris paribus*; simplicity is, as Kuhn (1977, pp. 320–329) put it, a theoretical virtue. Archer might respond that the advocate of the shifting view needs to appeal to other beliefs and propositional attitudes (or appropriate surrogates) to explain how and why shifting occurs. We grant this, and will say more about it in the next two sections. However, the kind of explanation we are concerned with, in raising this problem, concerns the SDS's behavior at a point in time. That's to say, we think the shifting account provides the basis for simpler answers as to why the SDS performs any particular action, as opposed to the question of why the SDS performs apparently inconsistent or incoherent actions over an extended period of time. (We don't take a position on which account is simpler than the other when it comes to explaining incoherent actions over an extended period.)

Second, to account for an SDS's behavior suggestive of a low degree of confidence in an undesirable proposition that $\sim p$, it appears we must attribute to the SDS beliefs with content like " p is more possible than $\sim p$ " or " p is more likely than $\sim p$ ". And when the SDS instead behaves as if she is almost fully convinced that $\sim p$, we instead appear to need to attribute beliefs with content like " $\sim p$ is much more possible than p " or " p is more likely than $\sim p$ ". Thus, a paradox akin to the static paradox threatens. In response, Archer might argue that when beliefs of the former kind are activated, beliefs of the latter kind are deactivated.

¹⁴ Note also that it may be possible to argue that a single degree of confidence in p is present, and that this functions as a degree of confidence in $\sim p$ when it is below a particular threshold.

See also footnote 8.

However, the consequence is that a kind of shifting in belief happens after all (although it's not a shifting concerning beliefs in p and not- p simpliciter). Furthermore, it appears that beliefs of the aforementioned kind need to function continuously during self-deception, on Archer's view, since she thinks that the doubts and suspicions (which entail such beliefs) continuously prevent the SDS from attaining belief in p or $\sim p$.

The strongest and final objection to Archer's account has already been covered in the prior discussion of Mele's account of self-deception. For an SDS to have a belief like "It is possible that p " or "It is likely that p " or " p is more possible than $\sim p$," as Archer suggests she must in order to suspect that p , the SDS must have a conception of possibility (and that it may come in degrees) or of probability (or some near surrogate). That is to say, provided it is impossible for one to fully believe in a proposition without grasping that proposition in its entirety.¹⁵ But it is possible for people who lack such concepts—such as children—to be self-deceived. Thus requiring such beliefs is inappropriate. Requiring that an SDS has a degree of belief (or degree of confidence) in a proposition, on the other hand, doesn't require that the SDS has the concept 'degree of belief', or even that the SDS is aware of her own degrees of belief.

¹⁵ Consider Dennett's (1969) example of a mother who teaches her young child that "Daddy is a doctor." If we ask the child if her father is a doctor, she will answer in the affirmative. But if we ask her to explain what a doctor is, she will not be able to answer satisfactorily. Thus we can at best say that the child "sort of" believes that her daddy is a doctor. (She probably believes "Daddy is known as a doctor", and possibly believes "Daddy's job name is 'doctor'".)

Nevertheless, the shifting view is consistent with some key elements of Archer's account of self-deception. For example, one might, in principle, endorse the shifting view while maintaining that "*anxiety* that p , and a *desire* to believe $\sim p$ " (Archer, 2013, p. 276) are present in all (paradigm) cases of self-deception. Conversely, however, one might not.¹⁶

3. Criticisms of the Shifting View

Before we demonstrate how the shifting view explains the inconsistent behavior of the SDS with reference to the example with which we began, we will briefly address some criticisms of the view.

As we noted previously, during our discussion of Mele's account of self-deception, Porcher (2012, pp. 77–78) criticizes the idea that self-deception involves shifting degrees of belief because the attribution of such "intermittent belief ... has absolutely no explanatory or predictive power". However, this criticism is originally due to Schwitzgebel (2010, p. 543), who uses the example of Juliet, a person who exhibits racist behavior in some contexts while sincerely declaring she is not a racist in others, in an illustrative capacity:

The main problem with this approach is that it leaves us without resources to describe the

¹⁶ Our concluding illustration, in Section 4.2, may be read with this in mind. We say that the husband fears that p (which might be equivalent, for some, to "is anxious that p ") and desires that $\sim p$ at different points. But we do not specify that the fear and the desire are continuously present. It might suffice instead, for example, for the husband to manifest fear that p over a threshold degree (which might even, on some views, be equivalent to hope that $\sim p$ over a threshold degree) in response to particular kinds of trigger.

subject's overall attitude. When Juliet is mowing the lawn, with general propositions about equality far from her mind, and neither seeing nor thinking about anyone of any skin color, what does she believe—that the races are intellectually equal or that they're not? The shifting approach leaves us stymied. It doesn't seem right to say she has no belief about the races at such a moment: Beliefs are dispositional; one doesn't cease believing when one falls asleep or turns one's mind to another topic; we don't need to find out what someone is currently doing to say what she believes the way we need to find out what someone is currently doing to say whether she is jogging. We need to be able to speak about Juliet globally, not just about her shifting judgments or assumptions in particular individual conditions. What's her general attitude?

We have already answered this charge, but a brief recap may be useful. The rhetorical question is based on two misconceptions: (1) degrees of belief are excessively changeable on the shifting view; and (2) beliefs are fully general attitudes. On (1), however, the shifting view doesn't imply that the subject's degree of belief in the relevant proposition *always*, or even *typically*, shifts when there are changes in the situation. Rather, the changes may be linked to specific triggers. So Juliet's degree of belief in "All races are equal in all respects" need not alter when she is "washing the dishes ... or sleeping peacefully" rather than mowing the lawn (Rowbottom, 2016, p. 733). In short, the shifting involved is not capricious. It occurs response to particular stimuli.

On (2), it's already normal to think that beliefs are subject to change in response to certain changes in context. For example, a subject who strongly believes that $\sim p$ may change her mind after consulting an expert on such matters who asserts that p . And we normally have no problem in predicting how someone's beliefs "are likely to change on a scenario-by-scenario

basis” in our daily lives (Rowbottom, 2016, p. 734). Thus belief isn’t thought of as a “general attitude”. Rather, it’s an attitude understood to apply to a subject “in some significant class of scenarios” (Rowbottom, 2016, p. 734). The shifting view only involves the idea that sometimes the class is narrower than others; it doesn’t involve a move from the general to the specific.

However, Schwitzgebel (2010, p. 543) continues by offering another, stronger, objection to the shifting view:

[I]t seems possible for Juliet in a single moment both to be having a racist reaction and to be sincerely judging that the races are intellectually equal—for example, when she’s having a racist reaction and trying to suppress it or when she’s grading a black student’s essay on intellectual equality, undervaluing the essay but regarding its conclusions as true. This is a possibility the shifting model gives us no means to accommodate.

Rowbottom’s (2016, p. 734) main response is that Juliet may be wrong about her own degree of belief in this case: “*it [only] seems to her that she thinks it’s true.*” He also defends the view that the degrees of belief might rapidly fluctuate in some circumstances, where appropriate triggers occur in quick succession. To this we may add that synchronic mismatches of such a kind don’t occur in the canonical examples of self-deception (or what Archer calls the paradigm cases). Thus even if this objection succeeds in showing that the shifting view doesn’t account for all cases where there’s a mismatch between assertion (or judgement) and action—as Rowbottom (2016) admits that it does not for independent

reasons—it doesn't show that shifting doesn't occur (or failing that, doesn't typically or often occur) in cases of self-deception.

4. Explaining Inconsistent Behavior by Appeal to Shifting

We now turn to how we can explain cases of self-deception where the SDS's behavior is inconsistent as a whole by appeal to shifts in degrees of belief. In order to provide a clear answer, we will first explain the two mechanisms by which a subject's degree of belief may shift.

4.1 The Two Forms of Shifting: Activation and Value Change

There are two forms of shifting: (i) “shifting active conditional degrees of belief,” and (ii) “shifting values of conditional degrees of belief” (Rowbottom, 2016, pp. 738–740). The first form of shifting may be referred to as “information-based” since it concerns the changes in the background information and evidence (Rowbottom, 2016, p. 739). In contrast, the second form of shifting may occur due to changes in non-doxastic attitudes.

To explain these two ways of shifting, we should begin by emphasizing the conditional nature of degrees of belief by analogy with conditional probabilities. Compare, for example, $P(\text{it will rain tomorrow}|\text{it has rained today})$ and $P(\text{it will rain tomorrow}|\text{it has not rained today})$; evidently, which of these is relevant for thinking about whether it will rain tomorrow depends on whether it has rained today or not. Similarly, if we represent degrees of belief as D-functions, $D(\text{it will rain tomorrow}|\text{it has rained today})$ will be “active” for an agent who assumes (or believes) it has rained today, whereas $D(\text{it will rain tomorrow}|\text{it has not rained$

today) will be “active” for an agent who assumes (or believes) it has not rained today—provided said agent grasps “it will rain tomorrow”, and so on.

Now the first kind of shifting—“information based shifting”—involves the activation of a conditional degree of belief (and the deactivation of another conditional degree of belief) due to a change in information. Consider the following example:

As I write ... I am confident that no one is standing outside my office. But if I hear a knock on my door—or even hear a noise that I mistake for a knock on my door—then I will be confident that someone is outside my office. (Rowbottom, 2016, p. 738)

Let p be “Someone is standing outside my office,” b be relevant background information (about how easy it is to approach the office without being heard from within, how frequently people attempt to approach the office without being heard from within, and so forth), and e be “The office door is being knocked on.”¹⁷ In the above example, the subject’s active degree of belief shifts from $D(p|b)$ to $D(p|b\&e)$. The value of $D(p|b)$ need not be altered in the deactivation process. If the person in the office changed his mind about e in the example above—e.g., because he realized the knock was too muffled to be on his own door, and was instead on the door of the neighboring office—his active degree of belief would revert to $D(p|b)$. (That is, assuming that whether someone is knocking on a different office door is not part of the *relevant* background information.) Similarly, $D(p|b\&e)$ need not have changed value before being activated, let alone come into being at the moment that $D(p|b)$ was

¹⁷ Background information is sometimes called “background knowledge,” but it need not be true. For more on this, see Rowbottom (2014) and Williamson (2015).

deactivated. In short, there may be a fact of the matter about how strongly one will believe something on the basis of potential changes of information, even if those changes never occur. For example, you'd believe that this paper had different authors if the names appearing on it were not our own (*ceteris paribus*).¹⁸

Second, however, the values of conditional degrees of belief may also change, perhaps in response to “a change in propositional attitudes other than beliefs, such as fears and desires” (Rowbottom, 2016, p. 739). For illustrative purposes, consider the following case presented by Schwitzgebel (2001, p. 78):¹⁹

In certain moods and in certain contexts, Antonio feels quite sure that the universe is guided by a benevolent deity. In other moods and contexts, he finds himself inclined to think of talk about God as “a beautiful metaphor” or even, sometimes, “a crock of hooey”. When his atheistic buddies at work mock religious belief, he does not join in, but neither does he feel an impulse to defend belief in God ... At the birth of a child or the death of a friend, he feels certain God is involved; when the church gossip group has invaded his house, the idea of taking literally talk about the existence of a benevolent deity strikes him as foolish.

¹⁸ A further possibility is that changes in occurrent beliefs—and not just dispositional beliefs—can affect which conditional degree of belief is active. This might explain cases where an SDS's behavior is affected by an existing belief being brought to (conscious) mind.

¹⁹ Originally, Schwitzgebel mentions this example in order to argue that there are “in-between believing” states that cannot be accounted for by appeal to degrees of belief. See Schwitzgebel (2001; 2010) for more details on in-between believing and Rowbottom (2007; 2016), Borgoni (2016), and Archer (2018) for criticisms.

As Rowbottom (2016) explains, changes in Antonio's state of information need not be responsible—and are plausibly not responsible—for changes in his degree of confidence in “God exists” in this scenario.²⁰ Antonio's conditional degrees of belief might instead change as his mood changes: “Antonio is a believer when in wonder, agnostic when calm, and a non-believer when irritated” (Rowbottom, 2016, p. 740). To this we might add that mood changes, and associated attitudes, may be understood to be a matter of degree in their own right, and that changes in degree of fear or hope, say, may be linked to changes in degree of belief in appropriate propositions. Consider, for instance, Mark the arachnophobe's degree of belief that “Domestic house spiders (or barn funnel weavers) are dangerous.” When reading about these spiders (*tegenaria domestica*) online after being reassured that they are harmless by a friend, knowing that no such spiders are in the vicinity, he is confident that there is nothing to fear from them. But when he later spots one in the bath, he experiences a mild sense of fear that causes him to doubt more strongly that this kind of spider is not dangerous. He summons his new housemate to remove it. Unfortunately for Mark, however, his housemate sees this as an opportunity to play a cruel prank. After catching the spider, he drops it down the back of Mark's shirt. Mark is terrified. He screams “Get it out! Get it out!” In that moment, Mark is confident that he is in imminent danger from the beast.

²⁰ This said, new information such as “There is a newborn baby in front of me” and “My friend recently died” *could* cause changes in Antonio's degree of belief that “God exists”. Some shifts are irrational.

It is important to note that the two forms of shifting might occur simultaneously.

Furthermore, changes might be causally related in some cases: “a change in degree of belief might trigger a change of mood, and vice versa” (Rowbottom, 2016, p. 740).

4.2 *Explaining Self-Deception by the Shifting View: A Brief Illustration*

We are now ready to show how it is possible to account for the case of a self-deceived husband who is betrayed by his wife, similar to that with which we began, in terms of shifting.

Let p be “My wife is having an affair”, and b be the relevant background information (concerning how people having affairs tend to behave, and so forth). The husband finds out that his wife might be having an affair after he overhears a telephone conversation that his wife is having. He learns that she is arranging to meet a man, and that she refers to him as “darling” (e). Hence, his active degree of belief that p suddenly shifts. It was

$D_{\text{husband}}(p|b)=0.01$ but is now $D_{\text{husband}}(p|b\&e)=0.7$. He is now keenly aware of how much he wants his wife to be loyal to him, and this provokes an emotional response. He feels despair, and his degree of belief changes to $D_{\text{husband-in-despair}}(p|b\&e)=0.8$. However, he then begins to look through his wedding photos and realizes how beautiful and special his wedding was (e_1).²¹ After this realisation, he feels somewhat relieved and rather happier. It seems like he must have misunderstood the significance of (e). His active degree of belief changes from $D_{\text{husband-in-despair}}(p|b\&e)=0.8$ to $D_{\text{husband-relieved/happier}}(p|b\&e\&e_1)=0.4$.

²¹ Or he might simply recall his belief that his wedding was special: this belief might become occurrent rather than dispositional, and this might affect his active degree of belief concerning p .

When he's later having dinner with friends, however, one friend suddenly declares "I don't want to tell you this, because I know it will devastate you. But I feel I have to. I saw your wife passionately kissing a man" (e_2). In that moment, $D_{\text{husband}}(p|b\&e\&e_1\&e_2)=0.9$ activates. The husband begins to feel a tinge of despair, but is motivated to challenge his friend's testimony by his desire for p to be false (and, *inter alia*, his belief that some testimony is false). Thus, he angrily blames his friend for doubting his wife's loyalty without conclusive evidence. His active degree of belief shifts once more to $D_{\text{husband-indignant}}(p|b\&e\&e_1\&e_2)=0.3$. He declares "You got the wrong person!" or "It was just normal social etiquette!" His friend then reluctantly backtracks to avoid conflict: "Okay! Maybe you are right. My fault." Down the husband's active degree of belief in p goes again, due to the change in information.²² And

²² The change here might be due, in part, to a false belief on the husband's part that his friend has been convinced that $\sim p$. Indeed, the shifting view of self-deception gels well with the idea that cognition is socially embedded, as hinted at previously, in so far as it allows for: (1) social triggers to effect changes in degrees of belief, and (2) social conditions to delimit the kinds of changes available (and hence interfere with self-deception). Thus the shifting view is consistent, for example, with the findings of Baumeister & Cairns (1992, p. 861) that:

[Exhibiting a] pattern associated with certain chronic tendencies toward self-deception ... repressors prefer to avoid threatening feedback when it is private. If a nonsalient threat is not noticed, perhaps it seems to disappear. Public circumstances, however, confer social reality on the ego threat ... One cannot expect others to ignore unflattering information about oneself, and so one must deal with it ... Under those circumstances, repressors apparently begin to attend closely to the evaluative feedback.

In short, the husband might be thought to feel the need to respond to the friend, rather than ignore him, because of the public nature of the accusation in the illustration. And a more

up again it later goes, after he comes back home and realizes, as he is about to enter the bedroom, that his wife is on the telephone. As a result of this—and his fear of encountering more evidence that his wife is being unfaithful, due to related fears about how he would then feel, and so forth—he coughs loudly before entering the room.²³

5. Conclusion

We have shown that explaining cases of self-deception by appeal to shifting degrees of belief avoids problems with two other recent revisionist account of self-deception, due to Mele and Archer. We have also shown that appeal to shifting degrees of belief—in conjunction with other propositional attitudes—is sufficient to explain a classic case of self-deception. We conclude that this approach should be taken much more seriously in future research on self-deception, as a live option for explaining many, if not all, cases of self-deception.

persistent friend, who continued to assert publicly that his wife was being unfaithful, might succeed in preventing the husband's self-deception from continuing.

²³ We do not intend to suggest that willful ignorance is a form of self-deception. We are sympathetic to the view of Lynch (2016, p. 552) that:

Willful ignorance and self-deception are associated with different ranges of doxastic states. Willful ignorance involves suspecting that *p*. Self-deception can involve suspecting that *p* too, but it can also involve believing that not-*p*, unlike willful ignorance.

On a degree of belief analysis of suspicion—as outlined in footnote 13—willful ignorance (so described) may also involve shifting degrees of belief, but the shifting involved will be narrower than that involved in self-deception. In short, the relevant degrees of belief will remain in the suspicion interval.

References

- Archer, S. (2013). Nondoxasticism about self-deception. *Dialectica*, 67, 265–282.
- Archer, S. (2018). Why ‘believes’ is not a vague predicate. *Philosophical Studies*, 175, 3029–3048.
- Asch, S. (1951). Effects of group pressure on modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men* (pp. 177–190). Pittsburgh, PA: Carnegie Press.
- Baumeister, R. F. & Cairns, K. J. (1992). Repression and self-presentation: when audiences interfere with self-deceptive strategies. *Journal of Personality and Social Psychology*, 62, 851–862.
- Borgoni, C. (2016). Dissonance and irrationality: A criticism of the in-between account of dissonance cases. *Pacific Philosophical Quarterly*, 97, 48–57.
- Davidson, D. (1982). Paradoxes of irrationality. In R. Wollheim & J. Hopkins (Eds.), *Philosophical essays on Freud* (pp. 289–305). Cambridge: Cambridge University Press.
- Davidson, D. (1985). Deception and division..In E. LePore & B. McLaughlin (Eds.), *Actions and events* (pp. 138–148). Oxford: Blackwell.
- Dennett, D. C. (1969). *Content and consciousness*. London: Routledge.

Galeotti, A. E. (2012). Self-deception: Intentional plan or mental event? *Humana.Mente: Journal of Philosophical Studies*, 20, 41–66.

Gendler, T. S. (2008a). Alief and belief. *Journal of Philosophy*, 105, 634–663.

Gendler, T. S. (2008b). Alief in action (and reaction). *Mind and Language*, 23, 552–585.

Gillies, D. (2000). *Philosophical theories of probability*. London: Routledge.

Kuhn, T. S. (1977). *The essential tension: Selected studies in scientific tradition and change*. Chicago: University of Chicago Press.

Leitgeb, H. (2014). The stability theory of belief. *The Philosophical Review*, 123, 131–171.

Levy, N. (2004). Self-deception and moral responsibility. *Ratio*, 17, 294–311.

Lynch, K. (2012). On the ‘tension’ inherent in self-deception. *Philosophical Psychology*, 25, 433–450.

Lynch, K. (2016). Willful ignorance and self-deception. *Philosophical Studies*, 173, 505–523.

Mele, A. (1999). Twisted self-deception. *Philosophical Psychology*, 12, 117–137.

Mele, A. (2001). *Self-deception unmasked*. Princeton: Princeton University Press.

Mele, A. (2012). When are we self-deceived? *Humana.Mente: Journal of Philosophical Studies*, 20, 1–16.

Pears, D. (1984). *Motivated irrationality*. Oxford: Oxford University Press.

Porcher, J. (2012). Against the deflationary account of self-deception. *Humana.Mente: Journal of Philosophical Studies*, 20, 67–84.

Rorty, A. (1988). The deceptive self: Liars, layers, and lairs. In B. McLaughlin & A. O. (Eds.), *Perspectives on self-deception* (pp. 11–28). Berkeley: University of California Press.

Rowbottom, D. P. (2007). “In between believing” and degrees of belief. *Teorema*, 26, 131–137.

Rowbottom, D. P. (2014). Information versus knowledge in confirmation theory. *Logique et Analyse*, 57, 137–149.

Rowbottom, D. P. (2015). *Probability*. Cambridge: Polity Press.

Rowbottom, D. P. (2016). How might degrees of belief shift? On action conflicting with professed beliefs. *Philosophical Psychology*, 29, 732–742.

Schwitzgebel, E. (2001). In-between believing, *Philosophical Quarterly*, 51, 76–82.

Schwitzgebel, E. (2010). Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, 91, 531–553.

Scott-Kakures, D. (2012). Can you succeed in intentionally deceiving yourself? *Humana.Mente: Journal of Philosophical Studies*, 20, 17–40.

Sorensen, R. (1985). Self-deception and scattered events. *Mind*, 94, 64–69.

Von Hippel, W. & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34, 1–56.

Williamson, J. (2015). Deliberation, judgement and the nature of evidence. *Economics and Philosophy*, 31, 27–65.