Les ateliers de l'éthique
The Ethics Forum

SELF-DECEPTIVE RESISTANCE TO SELF-KNOWLEDGE

Graham Hubbs

Volume 13, Number 2, Summer 2018

URI: https://id.erudit.org/iderudit/1059498ar
DOI: https://doi.org/10.7202/1059498ar

See table of contents

Publisher(s)

Centre de recherche en éthique (CRÉ)

ISSN

1718-9977 (digital)

Explore this journal

Cite this article

Hubbs, G. (2018). SELF-DECEPTIVE RESISTANCE TO SELF-KNOWLEDGE. *Les ateliers de l'éthique / The Ethics Forum*, *13*(2), 25–47. https://doi.org/10.7202/1059498ar

Article abstract

Philosophical accounts of self-deception have tended to focus on what is necessary for one to be in a state of self-deception or how one might arrive at such a state. Less attention has been paid to explaining why, so often, self-deceived individuals resist the proper explanation of their condition. This resistance may not be necessary for self-deception, but it is common enough to be a proper explanandum of any adequate account of the phenomenon. The goals of this essay are to analyze this resistance, to argue for its importance to theories of self-deception, and to offer a view of self-deception that adequately accounts for it. The view's key idea is that, in at least some familiar cases, self-deceived individuals maintain their condition by confusing a nonepistemic satisfaction they take in their self-deceived beliefs for the epistemic satisfaction that is characteristic of warranted beliefs. Appealing to this confusion can explain both why these self-deceived individuals maintain their unwarranted belief and why they resist the proper explanation of their condition. If successful, the essay will illuminate the nature of belief by examining the limits of the believable.

SELF-DECEPTIVE RESISTANCE TO SELF-KNOWLEDGE

GRAHAM HUBBS
ASSOCIATE PROFESSOR, UNIVERSITY OF IDAHO

ABSTRACT:

Philosophical accounts of self-deception have tended to focus on what is necessary for one to be in a state of self-deception or how one might arrive at such a state. Less attention has been paid to explaining why, so often, self-deceived individuals resist the proper explanation of their condition. This resistance may not be necessary for self-deception, but it is common enough to be a proper explanandum of any adequate account of the phenomenon. The goals of this essay are to analyze this resistance, to argue for its importance to theories of self-deception, and to offer a view of self-deception that adequately accounts for it. The view's key idea is that, in at least some familiar cases, self-deceived individuals maintain their condition by confusing a nonepistemic satisfaction they take in their self-deceived beliefs for the epistemic satisfaction that is characteristic of warranted beliefs. Appealing to this confusion can explain both why these self-deceived individuals maintain their unwarranted belief and why they resist the proper explanation of their condition. If successful, the essay will illuminate the nature of belief by examining the limits of the believable.

RÉSUMÉ :

Les explications philosophiques de l'auto-illusion ont eu tendance à mettre l'accent sur ce qui est nécessaire pour que quelqu'un soit considéré comme étant sous l'emprise de l'auto-illusion ou encore sur la façon dont quelqu'un parvient à un tel état. Moins d'efforts ont été dirigés vers les raisons pour lesquelles, si souvent, les individus sous l'emprise de l'auto-illusion opposent une résistance à l'explication véritable de leur condition. Cette résistance n'est peut-être pas essentielle à l'auto-illusion, mais elle est suffisamment courante pour constituer un explicandum approprié pour tout traitement adéquat du phénomène. Cet essai a pour buts d'analyser cette résistance, de défendre son importance pour les théories de l'auto-illusion, et de proposer une conception de l'auto-illusion qui en rend compte de manière adéquate. Cette conception repose sur l'idée suivante : au moins dans certains cas connus, les individus sous l'emprise de l'auto-illusion maintiennent leur condition en prenant la satisfaction non-épistémique qu'ils retirent de leurs croyances illusoires pour la satisfaction épistémique qui caractérise les croyances justifiées. C'est en faisant appel à cette confusion que l'on peut expliquer à la fois pourquoi ces individus conservent leur croyance infondée et pourquoi ils opposent une résistance à l'explication adéquate de leur condition. Si tant est qu'il y parvienne, cet essai éclairera la nature de la croyance en examinant les limites du croyable.

What a fool believes he sees
No wise man has the power to reason away
– Michael McDonald and Kenny Loggins,
"What a Fool Believes" (1978)

A fool who cannot be swayed by reason may not be self-deceived, but often he is. Consider the fool of McDonald and Loggins's song. He meets an old acquaintance who, he thinks, once longed for him romantically and might feel the same way again. She never had such feelings, and she never will; perhaps out of pity, she apologizes to the fool for the fact. Loggins and McDonald tell us that "as he rises to her apology, anybody else would surely know/ he's watching her go." Anybody else would surely know that she is going, and for good, because this is what the evidence overwhelmingly suggests. In spite of this evidence, the fool believes that she will return to him someday. On accounts of self-deception that Dion Scott-Kakures (2002) calls *deflationary*, little more needs to be said about the fool to depict him as self-deceived.[1] On Alfred Mele's version of deflationism, we need only to add that the fool's false belief results from him treating the relevant data in a motivationally biased way (Mele, 1997, p. 95). On Mark Johnston's version, we need only to add that the belief is the result of a distinct sort of mental tropism (Johnston, 1988, p. 86). On Annette Barnes's version, we must only add that what Johnston calls a tropism is specifically a form of anxiety avoidance and that the fool underestimates the causal impact that this has on the maintenance of his belief (Barnes, 1997, p. 117). These views are deflationary because they characterize the apparent goal-directedness of self-deception non-intentionally; they thus contrast with intentionalist views, such as Donald Davidson's (Davidson, 2004a; 2004b; 2004c). Scott-Kakures's own account takes a middle way between deflationism and intentionalism: he claims that in order to be self-deceived, the fool would have to engage in reflective reasoning that maintains his unwarranted belief (Scott-Kakures, 2002; 2009). On this view, the self-deceived fool would not intentionally bring himself into his condition, but nor would he fall into it as the result of some blind motivation.

Although these views disagree on the roles of intentions and reasoning in producing self-deception, they would all agree that the self-deceived fool believes that his acquaintance once had romantic feelings towards him and may feel this way again. Some recent accounts have called into question whether the self-deceived fool has any such belief. Eric Funkhouser and David Barrett would claim that, if the fool holds such a belief, then he is self-deluded but not self-deceived (Funkhouser, 2005; Funkhouser and Barrett, 2016; 2017). To be self-deceived, on their view, the fool would have to behave in a way that suggests he does not believe he has a chance romantically with his acquaintance. This behaviour is at odds with what the fool asserts—namely, that she might one day long romantically for him. To account for his nonlinguistic behavior, Funkhouser and Barrett would recommend that we ascribe to the fool the belief that she does not and will not care romantically for him; to account for his assertion to the contrary, we are to characterize him as falsely believing that he believes she

might want him one day. The difference here between self-delusion and self-deception is not merely terminological. Funkhouser explicitly asserts that self-deception is philosophically interesting in a way that self-delusion is not—in due course I will explain why. Jordi Fernández does not go as far as Funkhouser and Barrett, for Fernández allows cases they would describe as mere self-delusion to count as instances of self-deception (Fernández, 2013). Nevertheless, Fernández agrees with Funkhouser and Barrett that these cases are not of much philosophical interest; to be philosophically interesting, the fool must be characterized roughly in line with Funkhouser and Barrett's definition of self-deception. This requires that the fool hold the first-order belief that he has no romantic future with his acquaintance while simultaneously holding the false metabelief that he believes they will one day be together.[2]

Now it is one thing to explain what it takes for the fool to be self-deceived, which is a task pursued by all of the accounts just mentioned; it is another to explain why, so often, no one, no matter how wise, has the power to reason the self-deceived fool out of his condition. One might produce an intuitively plausible characterization of being self-deceived and perhaps even a causal account of how one arrives at the condition without explaining why, so often, self-deception resists rational revision. The resistance I have in mind may not be necessary for self-deception, but I believe it is common enough to be a proper explanandum of any adequate account of self-deception. The central goals of this essay are to analyze this resistance, to argue for its importance to theories of self-deception, and to offer a view of self-deception that adequately accounts for it.

To bring this resistance squarely into view, consider the following specific but straightforward way in which the wise man in the song might fail in his attempts to correct the self-deceived fool. Suppose the wise man points all of the relevant evidence out to the fool and asserts that, on its basis, it is reasonable to conclude that his acquaintance has never wanted him and will never want him. If the fool accepts this, then, at least for that moment, on any account of self-deception, he is no longer self-deceived. Assume this does not happen: the fool insists that, in spite of the evidence, his acquaintance wanted him once and may want him again. The wise man shakes his head and tells the fool that he says this only because admitting the opposite would be too painful. The wise man is wise, so he makes no assumptions about the fool's unconscious beliefs or higher-order beliefs; he simply notes and explains what the fool will not admit.

The fool can go one of two ways at this point. On the one hand, he can give in and accept the wise man's explanation. This would seem to force the fool to acknowledge that, as a first-order matter of fact, the acquaintance will never want him, but perhaps he can resist this; perhaps he can acknowledge that the evidence does not warrant what he claims about her, that he says what he does only because admitting otherwise would be too painful, yet still maintain that she has wanted him and might want him again. This case would be an instance of what some have called *epistemic akrasia*.[3] Like those who perform akratic actions, epistemically akratic individuals know they are being unreasonable; in

spite of this, they go on believing their unwarranted beliefs. We will consider this epistemic condition in several places over the course of the argument. At present, let us set it aside and suppose that the fool goes the other way and rejects the wise man's explanation. Without straightforwardly lying, the fool denies that there is anything he is failing to admit. The reason he says she has wanted him and might want him again has nothing to do with his desires, he asserts, though he grants that he does desire her affection. The reason he says these things, he claims, is that they are true. The fool, of course, is wrong, both about his acquaintance's feelings and about the reason he insists he knows what she feels. Call the former error *self-deceptive resistance to evidence*; this topic has been well examined in the literature on self-deception. The latter error, which I will call *self-deceptive resistance to self-knowledge*, has received less attention.[4] This resistance to self-knowledge constitutes a failure of self-explanation: by committing the error, the fool resists the proper explanation of why he claims (or, depending on one's view of self-deception, believes) that his acquaintance may one day want him again. This resistance is no accident, no simple mistake—as the wise man knows, it is part of an overall epistemic condition that enables the fool to go on believing what he does about his acquaintance.

The third section of this essay will provide an account of this self-deceptive resistance to self-knowledge. To work up to it, I will begin by discussing the views and arguments in Fernández (2009) and Funkhouser (2005).[5] Fernández might claim that my topic is not an interesting sort of self-deception, and Funkhouser might deny that it even counts as self-deception. My goal in responding to these complaints is to defend the claim that self-deceptive resistance to self-knowledge warrants "self-deceptive" as part of its label and to argue that this resistance is a proper explanandum for any account of self-deception. I will close this first section by showing that Funkhouser's and Fernández's accounts cannot explain the phenomenon. I will then turn to a discussion of deflationary and intentionalist views of self-deception. I will borrow from Scott-Kakures's discussion of deflationism and intentionalism, which I find insightful. Scott-Kakures's account improves upon these views, but in the end, I argue, it too lacks the resources to explain self-deceptive resistance to self-knowledge. The next section presents my account of the phenomenon. Its key idea—which, to the best of my knowledge, is novel to the literature on self-deception—is that self-deception involves a distinct sort of *confusion*. The sort of confusion I have in mind occurs when a person takes one thing to be something that it is not.[6] The fool, I will argue, confuses what I shall call the *epistemic satisfaction* of believing what is warranted with what I shall call the *thumotic satisfaction* that results from believing what he wants to be true. I will explain this distinction as well as my introduction of the term "thumotic" in section 3. At present, may it suffice to say the following: the fool thinks his belief is satisfying because he thinks it is warranted—i.e., he takes his satisfaction in the belief to be epistemic satisfaction—but he is confused, for the belief involves a different sort of satisfaction, the satisfaction one paradigmatically finds when one is esteemed or valued or admired in a way one wants, which I am calling thumotic satisfaction. My account here will draw upon recent work on the neurobiology of emotion by

Lisa Feldman Barrett (2017). I close the essay with some brief remarks on how the confusion account can explain so-called twisted cases of self-deception; this, I hope, will further elaborate the view.

The success of the confusion account turns ultimately on the accuracy of the characterization of the nature of belief in section 3. A key element of this characterization is the claim that the pleasure a belief might bring can partially determine whether or not one holds that belief, sometimes over and even against the belief's warrant. I think this is simply a fact, one that is readily demonstrated by the hopeful beliefs of some football supporters. Supporters may believe that this is the year for their team. The proper explanation for their holding this belief may in part be a matter of the evidence they consider—the season is young, and the team has not yet shown that it is bound to fail. This may, however, not be the whole story. To tell that story, one might also need to add that the hope brought by the belief is far more pleasant to the supporter than the doxastic alternatives, and that this too is part of the full explanation of the football supporters holding their belief. Because my account turns on a substantive view of some nonevidential but nonaccidental causes of belief maintenance, my discussion here may prove relevant to the reader who is unconcerned with self-deception but who is interested in doxastic voluntarism.[7] This essay will not investigate the possible connections between the confusion account and debates about believing at will. I mention the latter simply to flag a focus that it shares with the present discussion: the limits of the believable, as determined by the nature of belief.

1. AVOWAL AND CONFLICTING BEHAVIOUR IN SELF-DECEPTION

Let us start by considering the fool as we did at the outset—namely, as believing his acquaintance may one day want him again. As already noted, on Funkhouser's view this fool is self-deluded, not self-deceived; on Fernández's view this fool demonstrates only one of the two "remarkable features" of self-deception (Fernández, 2013, p. 381). The feature the fool demonstrates is what Fernández calls the *normativity of self-deception*, which is manifested by the fact that we are supposed to find his epistemic condition objectionable. I expect Funkhouser would agree that self-delusion involves some sort of normative failure. To be self-deceived, however, on Funkhouser's view, the fool would have to demonstrate the second of Fernández's features of self-deception—namely, the conflict feature.[8] There does not appear to be any conflict in the fool's overall psychological condition: he sincerely believes that his acquaintance may one day want him again. Both Funkhouser and Fernández think this case is importantly different from cases in which there is a conflict between a person's behaviour and what that person claims to believe. Consider, for example, Funkhouser's case of the self-deceived wife (Funkhouser, 2005, p. 302). She has ample evidence that her husband is having an affair with a friend. She claims to believe that he is faithful, but she behaves to the contrary; for example, she avoids driving by the friend's house at times when it seems likely that her husband might be there. Funkhouser thinks that the only way to make sense of the wife's driving behaviour is to attribute to her the belief that her husband is having an affair.

To make sense of her avowal to the contrary, we describe the wife as falsely believing that she believes her husband is faithful.

Adopting Funkhouser's labels for just this paragraph, we see that self-delusion and self-deception present different problems to explain. The problem with self-delusion is that evidence is not considered impartially, which leads to irrational belief formation. The problem with self-deception, by contrast, is that of squaring one's behaviour with one's conflicting avowals. There are at least two reasons one might find the latter problem more interesting than the former. Although Funkhouser and Fernández do not think of the problems this way, one might think the latter problem entails the former. If we treat the wife's avowal as an expression of what she believes, then not only does her behaviour fail to square with her avowal, but her belief also fails to square with the evidence, which includes her own behaviour. On this way of conceiving of the issue, self-deception is more problematic than self-delusion—it involves two problems, not just one—which might license finding the former more problematic than the latter. Funkhouser and Fernández do not characterize self-deception this way, however; again, they deny that the wife believes what she avows. I suspect this is due to their tacit acceptance of the following two principles of belief ascription. The first is that actions speak louder than words, so the tension between what the wife says and what she does is to be resolved by a belief ascription that explains her deeds, not her words. The second is a version of the principle of charity that prevents, as far as is possible, ascribing two flatly contradictory beliefs to the same person. They, it seems, take characterizing the wife's avowal without violating these principles to be more challenging than explaining irrational belief formation; this is the second possible reason one might find self-deception more interesting than self-delusion.[9]

I have no interest here in adjudicating what one should find more or less philosophically interesting, but I do want to argue, *contra* what is at least implied by Funkhouser's and Fernández's accounts, that there is something philosophically interesting about self-deceptive resistance to self-knowledge. First of all, even if resistance is not a necessary feature of self-deception, it is not uncommon for self-deceived individuals to be disposed to resist the proper explanation of their self-deceived condition. Funkhouser's cases suggest as much. For example, he describes the self-deceived wife as follows: "She laughs off the concerns of her girlfriends, and thinks to herself that Tony is certainly a faithful husband" (Funkhouser, 2005, p. 302). This demonstrates her self-deceptive resistance to the evidence, and her girlfriends lack the power to reason this resistance away. Now imagine one of these girlfriends saying, "Look, the only reason you insist Tony is faithful is that you can't bear to think otherwise, even though the evidence overwhelmingly suggests that he is cheating on you." It is surely possible, if not likely, that the self-deceived wife will dismiss this just as she laughed off their other concerns. This resistance to self-knowledge, then, is something that even a restricted view of self-deception such as Funkhouser's or Fernández's should seek to explain.

It is not clear, however, that they can explain the resistance without attributing a pair of contradictory beliefs to the wife. As we have already seen, they explicitly characterize her as believing that she believes that her husband is faithful. If she is not epistemically akratic—set this possibility aside—then she also must at least tacitly believe that her belief about what she believes is warranted. To believe the higher-order belief is unwarranted is simply to believe that she lacks the first-order belief; again, it is a central component of their view that she believes she has the favourable first-order belief about her husband. They are committed, however, to explaining her resistance to self-knowledge by characterizing her as also and simultaneously believing that her higher-order belief is unwarranted. This follows from the way they explain avoidance behaviour. They explain the wife's avoidance of driving by her friend's house by characterizing her as believing that her husband is cheating. Her resistance to self-knowledge also involves an object of avoidance: she is avoiding the fact that her belief that she believes her husband is faithful is unwarranted. According to their explanatory approach, she can avoid this fact only by believing that it is, indeed, a fact, so they are committed to characterizing her as believing that her higher-order belief about what she believes about her husband is unwarranted. If we are to avoid such contradictory ascriptions—and, again, this seems to be a principle that guides Funkhouser and Fernández—then this false-higher-order-belief approach to self-deception cannot characterize self-deceptive resistance to self-knowledge.[10]

2. INTENTIONALISM, DEFLATIONISM, AND REFLECTIVE REASONING

Funkhouser and Fernández are not alone in thinking that it is the internal psychological conflict of self-deception that makes the condition philosophically interesting. For example, in summarizing self-deception, Davidson says the following: "Finally, and it is this that makes self-deception a problem, the state that motivates self-deception and the state it produces coexist; in the strongest case, the belief that *p* not only causes a belief in the negation of *p*, but also sustains it" (Davidson, 2004c, p. 208). Unlike Funkhouser and Fernández, Davidson thinks that the conflict of self-deception exists between a pair of inconsistent first-order beliefs, one of which sustains the other. Nevertheless, all three agree that the internal psychological conflict of self-deception, whatever exactly it is, is the condition's most philosophically interesting element. Davidson's attempt to explain this conflict proceeds as follows. The self-deceived individual brings about an unwarranted belief through intentional activity, such as selectively attending to evidence that supports that belief. To explain how the unwarranted belief can be held even as the self-deceived individual goes on also holding its warranted contradictory, Davidson claims that the individual's mind is divided. Davidson insists that this division is functional;[11] he does this so as to allay worries of what we might call "homuncularism"—i.e., the idea that partitioning the mind requires partitioning the person into separate agents, each with its own agenda concerning what the person should believe.[12] The challenge of explaining the conflict of self-deception, then, is met by dividing the mind and attributing the contradictory beliefs to the separate parts.

Davidson's view is intentionalist because it claims that the self-deceived individual does something intentionally to bring about the belief that, though unwarranted, that individual prefers to hold. One need not divide the mind as Davidson does to hold an intentionalist view; indeed, Kent Bach holds an intentionalist view that agrees with Funkhouser, Barrett, and Fernández—namely, that the self-deceived individual does not believe the unwarranted but preferred belief (Bach, 1981).[13] Intentionalists provide a straightforward account of the way in which self-deceived individuals' motives can influence their beliefs: the motives produce intentional activities that allow the agents to ignore (if not eliminate) their unfavourable beliefs. A challenge for all intentionalists, then, is to explain how self-deceived individuals are able to do this while simultaneously acknowledging that the unfavourable beliefs are warranted by the evidence; as Scott-Kakures puts the point, the intentionalist is "under great pressure to claim that the contrary evidence is not *really* believed or that such evidence is somehow forgotten or otherwise pushed into inaccessibility" (Scott-Kakures, 2002, p. 577, italics in original). Scott-Kakures does not present the point as counting decisively against the intentionalist, but it provides a good reason for wondering whether one can account for self-deceived beliefs or avowals in less cognitively robust terms.

This is the project of the deflationist. There are, as noted at the outset, all sorts of deflationist views, but they all agree that self-deception is a sort of biased belief formation. Most views characterize this bias as motivated, but some deny that this is necessary (e.g., Patten, 2003). The challenge for these accounts, Scott-Kakures argues, is to distinguish self-deception from other sorts of biased belief formation, which do not seem to involve the characteristic tension of self-deception. The worry here is similar to the one that motivates Funkhouser and Barrett to distinguish self-delusion from self-deception.[14] For Scott-Kakures, however, the point is not that deflationist accounts pick out an uninteresting human condition; rather, it is that they cannot distinguish self-deception from the representational state a brute may enter under the effects of motivational bias. Scott-Kakures's example of such a brute is Bonnie the Cat, who is usually good at distinguishing cat-food sounds from non-cat-food sounds but who overreacts to non-cat-food sounds when she is very hungry. It seems natural to say that Bonnie's representations are affected in these situations by a motivational bias towards finding cat food; the challenge the deflationist faces is explaining how self-deception differs from Bonnie's biased representations.

The different limitations of intentionalism and deflationism are also shown by the difficulty each has in explaining self-deceptive resistance to self-knowledge. If the intentionalist is correct, then this resistance is intentional. For example, when the fool claims that he believes his acquaintance will return to him because this is what the evidence warrants, he is, on the intentionalist view, intentionally avoiding the correct explanation of his belief. Now the goal of avoiding a given sort of explanation is quite cognitively sophisticated; it is the sort of thing lawyers might do on behalf of their clients, and it is not something that Bonnie the Cat can so much as attempt. It is not clear how a person can intentionally

pursue such a sophisticated goal while sincerely denying doing so, but the intentionalist is committed to characterizing the self-deceived fool in this way, for such fools will sincerely deny the proper explanation of their resistance. The only hope for the intentionalist here, it seems, is the homuncularism that even Davidson wants to avoid. It is not clear that the deflationist fares any better, however; given the cognitive sophistication involved in avoiding the proper explanation of one's self-deceived condition, it is not clear how it could be pursued in anything less than an intentional manner. Deflationism may be an attractive alternative to intentionalism when it comes to explaining self-deceptive resistance to evidence, but its resources are strained if it is called on to explain the cognitively sophisticated resistance to self-knowledge.

Scott-Kakures's own view takes a middle path between intentionalism and deflationism.[15] According to Scott-Kakures, Bonnie the Cat is not self-deceived, because she plays no active role in generating or maintaining her biased representations. Self-deceived individuals, by contrast, play an active role in at least maintaining their unwarranted beliefs through reflective reasoning. Scott-Kakures grants that the unwarranted self-deceptive belief may be generated by non-self-deceptive means; what matters—and on this point, he is surely right—is the way in which the belief is maintained, not the way in which it is initially formed. On his view, for the fool in the song to be self-deceived, he must at least occasionally reflect on his beliefs about his acquaintance's feelings. When the fool does this, he must fail to reason clearly: his reasoning must be swayed towards believing what he prefers instead of what is warranted by the evidence.

Scott-Kakures explains this capacity for wrongful reasoning in terms of the psychological account of "pragmatic hypothesis testing."[16] According to this account, we are to make sense of the fool's reasoning by identifying the different costs associated with false-positive and false-negative hypotheses concerning his acquaintance. Consider the proposition, "My acquaintance has had and may again have feelings for me," conceived of as a hypothesis. If the fool settles in favour of this proposition and it is false, it is a false positive; if the fool settles against this proposition and it is true, then it is a false negative. From the fool's perspective, a false negative of this hypothesis would be worse than a false positive. If he does not believe it, but it turns out to be true, he will miss his chance with his acquaintance; if he believes it, but it turns out to be false, he will have given himself every chance with her, and he will have enjoyed temporarily believing in her affection.[17] This asymmetry of costs produces an asymmetry of acceptance thresholds for the positive and negative of the hypothesis: effectively, the fool tests the positive for sufficiency and the negative for necessity. He thus reasons his way to the false conclusion that he still has a chance with his acquaintance. Scott-Kakures is not the only writer to claim that reasoning plays a critical role in self-deception; he cites David Sanford (1988) as agreeing with him on this point. Sanford, however, depicts the reasoning involved as a false rationalization for one's belief, which serves to mask the genuine causes of one's self-deceived belief. Scott-Kakures argues that the reasoning conducted by the self-deceived individual functions not to disguise the genuine causes of the rele-

vant belief, but actually to maintain the belief; he says of the self-deceived individual that "her putative reasons are her reasons and they do, as a causal matter, explain why she has come to believe as she does" (Scott-Kakures, 2002, p. 597).[18]

It is surely correct that the reasoning performed by self-deceived individuals can sustain their unwarranted beliefs and thus play a causal role in the maintenance of these beliefs. It is not clear, however, that this can explain all of the beliefs that may be involved in self-deceptive resistance to self-knowledge. The putative reasons that constitute the fool's false self-explanation may indeed have the function of sustaining his self-deception. The existence of these self-explanatory beliefs, *contra* Scott-Kakures's account, is not explained by acts of reasoning that might have them as their conclusions. Their existence is explained by the fact that it would be too painful for the fool to admit the truth, either about what the evidence warrants believing about his acquaintance or about why he believes she will return to him one day. Scott-Kakures might think he can capture this fact by reiterating his point about acceptance thresholds, but appeal to these thresholds makes sense only if we think of the fool's false self-explanatory beliefs as something he might treat as hypotheses. This may not be impossible, but there are certainly cases (which, I suspect, are typical) in which the causal role is reversed. We can think of the fool's false self-explanation as comprising beliefs he has not adopted due to biased reasoning and does not sustain by such reasoning. For this fool, his self-beliefs do not have the status of hypotheses; they are not held on the basis of an examination of their epistemic credentials. The beliefs are prior to any reasoning about his acquaintance and they are prior to any reasoning about the cause of his beliefs about her. His false self-explanation expresses a presumption about the legitimacy of his reasoning, both about his acquaintance and about the cause of his beliefs about her. Commitment to such a false self-explanation is a condition of his reasoning counting as self-deceptive; it is not a product of any such reasoning.

If this is possible, then not even Scott-Kakures's account can fully explain self-deceptive resistance to self-knowledge. For Scott-Kakures, this resistance would have to manifest itself as an openness to the possibility of the truth of the proper self-explanation, which is then resisted. If the fool is as I have described him, the object of resistance is the possibility that the proper self-explanation is true—again, his preferred self-explanation is taken as fact, not treated as a hypothesis. By focusing on the role that reasoning can play in self-deception, however, Scott-Kakures's account points us in the right direction. The key is to see how we can become confused about what governs our mental activity as we rationalize our beliefs and actions, for it is in terms of this confusion, I think, that we are able to explain self-deceptive resistance to self-knowledge. I turn now to this.

3. CONFUSION AND RESISTANCE

My account starts with what I take to be an uncontroversial point: reasoning is a sort of agential mental activity.[19] As such, reasoning shares some metaphysical characteristics with other sorts of agential activity, including intentional

bodily action. Our agential activity, both mental and bodily, is a product of our nature as rational living beings. As specifically *living* beings, we have a general tendency to do what is satisfying and not to do what is dissatisfying. This tendency can affect the body by leading us to indulge in sexual, gastronomic, or drug-induced pleasure or to avoid arduous or fearful situations. This tendency can also affect the mind, and in similar ways—for example, in the same way that we have a tendency against putting ourselves in fearful or sad situations, we have a tendency against thinking fearful or sad thoughts. I am not denying that we sometimes tend towards or even fixate on unpleasant thoughts—more on this in section 4. At present, I am simply noting something that I think Johnston's tropistic account and Barnes's anxiety-avoidance account get right: there is a tendency of the mind to avoid thinking unpleasant thoughts.[20]

According to the constructivist account of emotion advanced by Lisa Feldman Barrett (2017, ch. 4), the basic affective components of all of these tendencies are surprisingly simple. They are the product of introception and are two-dimensional: they all are valenced and so are more or less pleasant, and they all involve some level of arousal, the lowest being exemplified by the lethargy characteristic of deep depression. On her view, the affective components of anger and fear are the same—both are more unpleasant than pleasant and involve arousal rather than calmness or lethargy. The difference in subjective experience between these two emotional categories is primarily a matter of nonaffective interpretation, which is heavily influenced by a person's understanding of emotional concepts. Affect can thus be common across emotional categories; it can also be common across perceptions of fact and mere contemplation of thoughts. On her account, the affect underlying severe ophidiophobes' aversion upon seeing actual snakes has the same components—displeasure and arousal—as their aversion to merely thinking about snakes. Combining this line of thought with that of the previous paragraph yields the following: there is a tendency, in many if not most of us, to avoid unpleasant and arousing affect, whether that affect is part of one's body's engagement with the external world (including, e.g., actual snakes) or is merely part of one's mind (including, e.g., mere thoughts of snakes). This holds for other agential tendencies as well (e.g., towards pleasure).

I take the following also to be uncontroversial: many if not most of us have a tendency to want the approval of at least some humans, whose opinion of us we value in some way. I call the satisfaction at which this tendency aims *thumotic*. The root of this term is the Ancient Greek "*thumos*," which is commonly translated into English as "spirit" and is sometimes understood narrowly as a psychological faculty of anger or, still more narrowly, revenge. Not hamstrung by any English version of the term, I intend "thumotic" more broadly, in line with the variety of uses of "*thumos*" in Ancient Greek—I use it to pick out the sorts of self-satisfaction that follow from one's perception of the positive judgments of others and the sorts of dissatisfaction that follow from corresponding negative judgments.[21] The satisfaction of many sorts of pride is, in my sense, thumotic. When one feels good in victory, the satisfaction is thumotic; when one basks in the praise of a superior, the satisfaction is thumotic. These are cases in which one

feels pleasant affect at having achieved a particular social standing—the good feeling follows from how one perceives oneself to be judged by others. Various sorts of admiration can also give rise to thumotic satisfaction: teachers may take thumotic satisfaction in the admiration of their students; someone fit and smartly dressed may take thumotic satisfaction in the desirous glances of others. The affect in all of these cases is pleasant; it seems typically to be arousing as well. Shame, guilt, and other emotions of social inadequacy or social failure involve thumotic dissatisfaction. The affect of these is unpleasant; when they manifest as a sort of sadness or depression, they are also low on arousal, not stimulating.

These remarks on thumotic satisfaction and dissatisfaction are, admittedly, quite rough—one might worry that these are ad hoc categories, and one might wonder what else belongs to them outside of the examples I have just listed. These are worthy concerns, which I will not address here. It will presently suffice if I can adequately distinguish the satisfaction one can take from the esteem of others from the satisfaction one can take in learning or discovering or understanding something, which I call *epistemic* satisfaction. The latter may seem like a strange sort of satisfaction, but it is found, I believe, in a variety of mundane mental activities. Consider the philosopher who wonders how to explain why we cannot form beliefs at will. The perplexity of the person troubled by this topic is accompanied by a distinct sort of epistemic tension, which that person seeks to resolve with a satisfying solution. Or consider the person who loves detective stories; this person takes pleasure in the explanatory tension posed by a good mystery and seeks to resolve this tension with a satisfying resolution to the story. In both of these cases, an explanatory tension persists until an answer is found that satisfactorily resolves the tension. The satisfaction in each case is epistemic, for it is satisfaction at having arrived at what one takes to be an understanding of the matter at hand. In these cases, the affect is pleasant and calming, for it resolves the arousal of the epistemic tension. To be clear, I am not claiming that all, or even most, epistemically satisfying states involve a perceived *feeling* of satisfaction. The satisfaction in question is often experienced as a sort of epistemic *ataraxia* or tranquility—many, perhaps most, epistemically satisfying states are those from which a feeling of epistemic dissatisfaction is absent.[22] The dissatisfaction in question is the genus to which the negative affect of cognitive dissonance belongs—satisfaction can simply be the absence of this feeling of negative affect.[23] Although it is easiest to exemplify epistemic satisfaction by focusing on cases in which a felt perplexity is resolved, no such feeling is necessary for a belief or explanation to be epistemically satisfying.[24]

If the difference between thumotic and epistemic satisfaction is not already clear, note that people sometimes seek to satisfy explanatory tensions even though they believe the resolution they seek might be otherwise unpleasant. Consider the person who is betrayed by a friend and who wants, among other things, an explanation that makes sense of the friend's betrayal. It seems wrong to think that this person aims at an overall condition of pleasure in wanting to understand the cause of the friend's betrayal, yet the betrayed person still wants an explanation that makes what the friend has done intelligible. The betrayed person wants an

explanation that is epistemically satisfying, even if it is one that also involves a feeling of disappointment, a sense of being insufficiently valued by one's friend. This latter feeling is one of thumotic dissatisfaction. Indeed, it may involve anger, perhaps at the friend for the betrayal, or perhaps at oneself for trusting someone who turns out to be untrustworthy. Even if the feeling is not the unpleasant arousal of anger, the dissatisfaction is thumotic: it is the displeasure, whatever the level of arousal, that follows from the negative evaluation implied by the friend's act of betrayal. The betrayed person may thus pursue an explanation that is epistemically satisfying even while expecting it will be thumotically dissatisfying.

Although we are capable of distinguishing epistemic satisfaction from thumotic satisfaction, we do not always exercise the capability. On occasion, the failure to exercise the capability can lead to a distinct sort of confusion, which in turn, I argue, can produce self-deception. I should be clear here about what I mean by "confusion" and its cognates. As stated in the introduction, as I understand the phenomenon, confusion occurs when a person takes one thing to be something that it is not. One may confuse instances of a common kind (as happens when, e.g., I confuse a person with that person's twin), or one may confuse instances of different kinds (as happens when, e.g., I confuse molybdenum with aluminum). One may be confused because one lacks the ability to discriminate between two discriminable items, but one may also be confused because one has a discriminative ability yet fails to utilize it properly on some occasion. One may be confused without knowing that one is confused, so confusion should be distinguished from the affective condition of feeling perplexed. If, for example, I do not know there is anything called "molybdenum," then I will not know when I confuse molybdenum with aluminum that I am confusing the two. Even if I do know that they are two distinct sorts of metals, I may on occasion confuse the one with the other without being aware in the least that I am doing so.

Many cases of self-deception, I think, result from an individual confusing thumotic satisfaction with epistemic satisfaction. These sorts of satisfaction can be confused when their underlying affective components are sufficiently similar. Consider the fool again. The thought that his acquaintance may return to him one day satisfies him. The valence of this thought is pleasant, and the valence of the alternative is unpleasant. The satisfaction he finds in this thought cannot be exclusively epistemic, for an impartial review of the evidence would lead anyone, himself included, to conclude that she never has wanted him and never will want him. To make the point clear, assume that, if the fool were given similar evidence about a different pair of people, he would immediately conclude that the admired individual has no reciprocal affection for her admirer. At least part of the satisfaction he finds in the thought is thumotic—it is the pleasure he takes in believing that she wants him. He takes the relevant pleasure, however, as an indication that the thought *is* true, not that it is what he *wants* to be true. The thought feels right to him: he confuses this feeling, which is the pleasure of thumotic satisfaction, with the affect of epistemic satisfaction. So confused, he believes the thought. (This is not to say that the thought must be completely devoid of epistemic satisfaction: more on this below.)[25]

Someone sympathetic to Scott-Kakures might complain here that the confusion account, at least as it has just been presented, does not capture the active role the self-deceived agent plays in maintaining his or her condition. In order for the confusion account to be adequate, the complaint here goes, it must depict self-deceived individuals as playing some agential role in maintaining their confusion, as that, according to the account, is the source of their self-deception. This complaint can be met, I think, if the view is augmented by adding the following conditional claim: if self-deceived individuals are asked why they hold their unwarranted beliefs, they will, without lying, typically provide epistemic reasons in support of them, and they will never acknowledge that the unwarranted beliefs are unwarranted. Were such individuals epistemically akratic, they would not offer any epistemic defence but instead would admit that their beliefs are unwarranted; as ever, let us set this possibility aside. Consider again the fool from Loggins and McDonald's song: because he is confused about the satisfaction he takes in his belief, he will not say that he holds it because it would be too painful to do otherwise; instead, he will offer reasons that he takes to warrant holding the belief. His disposition to defend and to support his self-belief is clearly agential—this is the sort of thing agents and agents alone do—so his condition is maintained, at least counterfactually, by his agency.

Appealing to confusion and counterfactuals as I have gives rise to another, perhaps more basic, worry. Most, if not all, cases of self-deception at least appear to be motivated. Confusion, however, is often the result of an accident; how then can my account capture the at least apparently motivational aspect of self-deception? To answer this worry, consider the order of explanation I have been presenting: the self-deceived individual finds some thought satisfying, he confusedly takes the satisfaction to be epistemic satisfaction when it is in fact thumotic, so he believes the thought, and is thereby disposed to explain the belief as he would other beliefs of his. It should be clear that the belief here is motivated; it is held not on the basis of epistemic credentials but instead for the thumotic satisfaction it brings. The question, then, is whether the confusion itself is motivated, or whether it is a mere accident, or whether there is some third way it might come about.

To understand self-deception as self-*deception*, I think we must understand the confusion as motivated. This, however, does not require that we see it as some strategy executed by the self-deceived individual. For a given individual's thought to be a self-deceived belief, the thought, obviously, must be a possible belief. As such, it is the sort of thought that can be epistemically satisfying. It also, however, must be the sort of thought that can be thumotically satisfying to the individual; were it not, the person's epistemic stance towards the thought would be simply and exclusively determined by the extent to which that person finds it epistemically satisfying. (Perhaps most of our beliefs are like this, devoid of any thumotic component, and perhaps this is why most of our beliefs are not candidates for self-deception.) If the thumotic satisfaction of such a thought can be maintained only by believing it, and if the affective consequence of giving up the thought would be intolerably dissatisfying, then nothing more is needed to

generate self-deceived confusion. The presence of these elements does not guarantee self-deception; it is a mark of epistemic courage to believe what is warranted even when the belief is thumotically or otherwise dissatisfying. To exercise this courage, however, one must clearly distinguish the different sorts of satisfaction that are present in a given belief. If one is not skilled at distinguishing these elements, then it is easy for one to settle on what is thumotically satisfying. Should this happen, then the resulting condition—including the confusion that sustains it—is motivated, for the result is that one believes what one wants.[26]

Describing the self-deceived individual as believing what is thumotically satisfying might give rise to yet another basic worry. One might complain that, although some thoughts are thumotically satisfying and others are not, only someone pathological would take the thumotic pleasure of some thought as an indication of its being true. Most take self-deception, however, to be a nonpathological sort of irrationality. If the confusion account succeeds only by making all self-deception pathological, the complaint concludes, then it is not plausible. The way to resist this complaint, I think, is to note that confusion-based cases of self-deception often (if not always) involve a genuine element of epistemic satisfaction in the overall condition. Unlike confusing, for example, aluminum with molybdenum, which involves the complete confusion of the one with the other, there is likely to be an element of epistemic satisfaction sustaining the fool's belief about his acquaintance. If he concocts a rationalization for the belief, that rationalization will supply reasons for it that, were they true and decisive, would warrant his holding it. The rationalization, like any chain of reasoning a person finds compelling, is epistemically satisfying. Indeed, this explains why the fool seizes on these reasons as a rationalization for his belief; because he takes the belief to be epistemically satisfying, and because the relevant reasons provide grounds for this epistemic satisfaction, he settles on them as his explanation for his belief. It is not pathological to hold a belief that one takes to be backed by epistemically satisfying reasons, so the fool's condition, although irrational, is not pathological.

These remarks concerning self-deceived rationalization bring the confusion account partially in line with Scott-Kakures's view. The views differ, however, on the way in which they explain self-deceptive resistance to self-knowledge. Specifically, they differ on how to account for the fool's false self-explanation of why he believes his acquaintance will return to him one day. As noted in the previous section, Scott-Kakures seems committed to claiming that biased hypothesis testing is the cause of the fool's belief that his belief about his acquaintance is evidentially grounded.[27] The fool's self-belief here, however, is not the result of hypothesis testing; rather, it expresses the presumption that his belief that his acquaintance will return to him one day is warranted. This self-belief, which concerns the proper explanation for his maintaining his belief about his acquaintance, may be only tacit. At least as we have been conceiving of the case, however, it must be there; it explains why he resists the wise man's (correct) explanation of his condition. This self-belief is an immediate result of

the epistemic satisfaction he mistakenly takes in his belief about his acquaintance: the fool believes the self-belief to be true because its truth is a condition of the truth of his belief about his acquaintance. As long as he takes his satisfaction in the latter to be, at base, epistemic, he will also take the former to be true. Scott-Kakures is right, I think, to insist that an adequate account of self-deception must explain the agent's role in maintaining the condition in non-intentional terms; I also think he is right to focus on reasoning as the primary means of this maintenance. His hypothesis-testing model, however, cannot correctly characterize the self-deceived individual's resistance to being told either that or why he is maintaining his self-deception. If a disposition to resist the proper explanation of one's self-deceived belief is a common feature of the condition—and I hope to have shown here that it is—then the confusion account should be preferred for its ability to explain it.

4. CONCLUSION: TWISTED SELF-DECEPTION

I have argued that the confusion account can explain both what the fool believes and why no wise man has the power to reason him out of his condition. The fool's anticipation of the pain he would feel upon giving up his preferred belief drives both his maintenance of the belief and his insistence that it is warranted. Not all cases of self-deception are like the fool's, however; there are cases of self-deception in which the individual maintains an unwarranted belief whose valence, at least apparently, is unpleasant. Mele (1997, 1999, 2001) has dubbed this "twisted" self-deception. For an example, imagine a jealous husband who has no evidence that his wife is cheating on him and overwhelming evidence that she is faithful, yet who persists in believing that she is having an affair. This sort of case poses at least a prima facie challenge to views such as Mele's and Barnes's, for it is not immediately obvious what could motivate this husband to maintain his jealous belief, nor is it immediately obvious how this husband's condition reduces anxiety (indeed, it might seem to provoke it). Both Mele (1997, 2001, ch. 5) and Barnes (1997, p. 44-46) have sought to defend their views from this challenge. Funkhouser (2005, p. 307-309), Scott-Kakures (2009, p. 101-105), and Fernández (2013, p. 386) also consider twisted self-deception; for them, it is a phenomenon that a full account of self-deception should be able to explain.[28] The confusion account can, I think, explain the phenomenon, and showing how it does will help to elaborate the view, in part by saying a bit more about thumotic satisfaction. Let me close, then, with some remarks on how it may be used to approach twisted cases.

In all cases, twisted and nontwisted alike, the explanatory value of the confusion account comes out when considering resistance to self-knowledge. Imagine, then, that the jealous husband is correctly told that he does not think his wife is unfaithful because that is what his best estimate of the evidence suggests; rather, he believes what he does because, even though the belief is unpleasant, it is somehow satisfying to him. Suppose he resists, asking how such an unpleasant belief could be satisfying in any way. Here, we can answer him by first taking a clue from etymology. The term "satisfaction" derives from Latin terms that

signify the idea of doing ("*facere*") enough ("*satis*"). There are all sorts of things that we can satisfy by doing enough: we can satisfy demands, expectations, contractual obligations, etc. These sorts of satisfaction need not involve pleasure. The satisfaction the jealous husband takes in his unpleasant belief, then, could be the thumotic satisfaction of angrily upholding a code of honour, which he confuses with willingly accepting an unpleasant truth. These are confusable, for both anger and considering unpleasant truths are negatively valenced. To be sure, there may be other ways of characterizing the jealous husband; all I want here is to sketch the kind of explanation the confusion account is positioned to give. It does not have to appeal to pleasure to make sense of satisfaction, so nor must it appeal to pleasure to make sense of confusion, self-deception, or resistance to self-knowledge. Reflecting on twisted cases, then, may help not only clarify the confusion account but also bring out its explanatory power.

ACKNOWLEDGMENTS

The ideas in this paper have been developing for well over a decade—gratitude for helping to shape them goes to more people than I can now remember. I would like to explicitly thank Joe Camp, Matthew Chrisman, Peter Machamer, John McDowell, Sebastian Rödl, and Kieran Setiya for help at the beginning of this project. For help finishing it, I thank Todd Nagel, as well as the editors and referees of this journal.

NOTES

[1] Scott-Kakures adopts this term from Alfred Mele's self-characterization (Scott-Kakures, p. 577, n. 3) but applies it more broadly than Mele does.

[2] Funkhouser, Barrett, and Fernández are not the first to develop views along these lines. They all acknowledge the influence of Robert Audi (1985, 1988, 1997) and Kent Bach (1981, 1997) on their work.

[3] See, e.g., Hookway (2001), Owens (2002), and Greco (2014).

[4] Although little has been written about this resistance to self-knowledge, plenty has been written on self-deception as involving a failure of self-knowledge. Scott-Kakures, Funkhouser, and Fernández all present self-deception in this way, as do Sanford (1988), Cohen (1992), Holton (2001), and Bilgrami (2006).

[5] I focus on Funkhouser (2005) rather than Funkhouser and Barrett (2016) because the former relates more immediately to the present discussion.

[6] Joseph Camp (2002) has suggested that the sort of confusion I am discussing might be more perspicuously labeled "ontological confusion" (Camp, 2002, p. 3). I, like Camp, will stick to using the simpler "confusion." My thinking about confusion owes much to Camp's insightful work on the topic.

[7] The literature on this topic is vast. See, *inter alia*, Alston (1988), Audi (2001), Bennett (1990), Chrisman (2008), Hieronymi (2006), Setiya (2008), Scott-Kakures (2000), Shah and Velleman (2005), and Williams (1970).

[8] See Lynch (2012) for an extensive discussion of this feature.

[9] If this last point is what leads them to find self-deception interesting, then they are not alone. Tamar Gendler (2010b, 2010c) has introduced the concept of "alief" to account for cases in which individuals act contrary to what they believe; one of her main explanatory goals is to account for these actions without characterizing the individuals as holding inconsistent pairs of beliefs. For critiques of this notion of alief, see Hubbs (2013) and Mandelbaum (2013). Gendler (2010a) does not explain self-deception in terms of aliefs; instead, the phenomenon is characterized as a sort of pretense. Whatever virtues this account might have, it will not help explain the fool's self-deceptive resistance to self-knowledge—sustaining this resistance is not a matter of pretending.

[10] For more problems with Funkhouser's position as it is elaborated in Funkhouser and Barrett (2016), see Doody (2017); for a response, see Funkhouser and Barrett (2017).

[11] On this, see Davidson (2004a, p. 185).

[12] I take the term "homuncularism" from Johnston (1988).

[13] Other, somewhat more recent intentionalist views can be found in Talbott (1995) and in Bermúdez (1997, 2000).

[14] Similar, but not identical. Funkhouser and Barrett claim that "philosophers and psychologists have a hard time keeping the deception in self-deception" (Funkhouser and Barrett, 2017, p. 682). The deception mentioned here is distinguishable from the tension discussed above.

[15] Scott-Kakures (2002), elaborated and developed in Scott-Kakures (2009).

[16] Scott-Kakures is not alone here; see also Mele (2001), as well as Scott-Kakures (1996), which draws on and critiques Mele (1987). As Scott Kakures (2009, p. 76, n. 12) notes, sources for this approach include Friedrich (1993) and Trope and Liberman (1996).

[17] With something close to this last point in mind, McDonald and Loggins tell us, "What seems to be is always better than nothing."

[18] Scott-Kakures (2009) elaborates this view using the resources of cognitive dissonance theory—these elaborations are irrelevant to the criticism I pursue in this section (but cf. n. 23 n. 27).

[19] For more on agential mental activity, see Soteriou (2005) and Soteriou and O'Brien (2009).

[20] This is not a recent discovery: the fact that our minds are susceptible to forces that are standardly thought of as operating on the body is a major theme of Freud's work. For a particularly illuminating discussion of the matter, see Freud (1911).

[21] See Padel (1992, p. 27-30) on "*thumos*." Drawing very loosely on the view Socrates presents in Plato's *Republic*, I take thumotic satisfaction to be distinguishable from the appetitive "pleasures of food, drink, sex, and others that are closely akin to them" (Plato, 1992, p. 111) and from the epistemic satisfaction I discuss later in this section. Mine can only be a loose interpretation, however, as Socrates characterizes *thumos* as the part of the soul we "get angry with" (Plato, 1992, p. 111). I take the space between food, drink, and sex, on the one hand, and learning and knowledge, on the other, to leave room for social emotions other than anger.

[22] Barrett agrees that an affective condition can be satisfying without involving a detectable feeling; "even a completely neutral feeling is affect" (Barrett, 2017, p. 72).

[23] Scott-Kakures (2009) draws explicitly on the literature on cognitive dissonance to develop an account of self-deception. I return to this below; cf. n. 27.

[24] Jonathan Lear (1988) discusses this tendency towards epistemic satisfaction as a desire for understanding; adapting a term from Melanie Klein, he calls this desire *epistemophilia* (see Lear, 1988, p. 3-10). I might have characterized this tendency as aiming at truth or knowledge or understanding, but I wish to avoid the debates that surround these topics.

[25] I have focused here just on valence, but for the two sorts of satisfaction to be confusable, the levels of arousal will also need to be sufficiently similar. I suspect there are some ways of depicting the fool where the arousal is elevated and others where it is low. I set this aside. My point here is made, I hope, by understanding how the common valence—which in the fool's case is pleasure—could be confused.

[26] By using the virtue-theoretic language of "epistemic courage," we can accurately locate, I think, the sort of responsibility that self-deceived individuals have for their condition. In saying this, I thus disagree with Neil Levy's view that self-deception is a simple mistake that lacks any necessary connection to culpability (Levy, 2004). A full discussion of the normative questions surrounding self-deception is beyond the scope of the present essay. For more on the topic, see Mary van Loon's contribution to this issue of *Les ateliers de l'éthique/The Ethics Forum.*

[27] Also, as noted in n. 23, Scott-Kakures has elaborated this view by drawing on cognitive dissonance theory. This is an account of what might drive one to test a (biased) hypothesis, but it keeps the hypothesis-test model intact. Moreover, cognitive dissonance is necessarily a dissonance *between* two separate epistemic states. The confusion view on offer here concerns two sorts of satisfaction one might find in a *single* epistemic state.

[28] For other discussions on twisted self-deception, see Nelkin (2002) and Michel and Newen (2010).

REFERENCES

Alston, William, "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives*, vol. 2, p. 257–299, 1988.

Audi, Robert., "Self-Deception and Rationality," in Martin, Mike (ed.), *Self-deception and Self-understanding*, Lawrence, University of Kansas Press, 1985, p. 169-94.

———, "Self-Deception, Rationalization, and Reasons for Acting," in McLaughlin, Brian and Rorty, Amélie Oksenberg (eds.), *Perspectives on Self-Deception*, Berkeley, University of California Press, 1988, p. 92-120.

———, "Self-Deception vs. Self-Caused Deception: A Comment on Professor Mele," *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 104.

———, "Doxastic Voluntarism and the Ethics of Belief," in Steup, Matthias (ed.), *Knowledge, Truth, and Duty*, New York, Oxford University Press, 2001, p. 93-114.

Bach, Kent, "An Analysis of Self-Deception," *Philosophy and Phenomenological Research*, vol. 41, no. 3, 1981, p. 351–70.

———, "Thinking and Believing in Self-Deception," *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 105.

Barnes, Annette, *Seeing through Self-Deception*, Cambridge, Cambridge University Press, 1997.

Barrett, Lisa Feldman, *How Emotions Are Made*, New York, Mariner, 2017.

Bennett, Jonathan, "Why Is Belief Involuntary?" *Analysis*, vol. 50, no. 2, 1990, p. 87–107.

Bermúdez, José, "Defending Intentionalist Accounts of Self-Deception," *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 107-108.

———, "Self-Deception, Intentions, and Contradictory Beliefs," *Analysis*, vol. 60, no. 4, 2000, p. 309-319.

Bilgrami, Akeel, *Self-Knowledge and Resentment*, Cambridge, Harvard University Press, 2006.

Camp, Joseph, *Confusion: A Study in the Theory of Knowledge*, Cambridge, Harvard University Press, 2002.

Chrisman, Matthew, "Ought to Believe," *Journal of Philosophy*, vol. 105, no. 7, 2008, p. 346-370.

Cohen, L. Jonathan, *An Essay on Belief and Acceptance*, Oxford, Clarendon Press, 1992.

Davidson, Donald, "Paradoxes of Irrationality," in Donald Davidson, *Problems of Rationality*, Oxford, Clarendon Press, 2004, p. 169-188.

———, "Incoherence and Irrationality," in Donald Davidson, *Problems of Rationality*, Oxford, Clarendon Press, 2004, p. 189-198

———, "Deception and Division", in Donald Davidson, *Problems of Rationality*, Oxford, Clarendon Press, 2004, p. 199-212.

Doody, Paul, "Is There Evidence of Robust, Unconscious Self-Deception? A Reply to Funkhouser and Barrett," *Philosophical Psychology*, vol. 30, no. 5, 2017, p. 657-676.

Freud, Sigmund, "Formulations on the Two Principles of Mental Functioning," in Stratchey, James and Anna Freud, (collaborating trans.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol. 12, London, Hogarth Press, 1911/1966, p. 218-226.

Fernández, Jordi, "Self-Deception and Self-Knowledge," *Philosophical Studies*, vol. 162, no. 2, 2013, p. 379–400.

Friedrich, James, "Primary Error Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena," *Psychological Review*, vol. 100, no. 2, 1993, p. 298-319.

Funkhouser, Eric, "Do the Self-Deceived Get What They Want?", *Pacific Philosophical Quarterly*, vol. 86, no. 3, 2005, p. 295–312.

Funkhouser, Eric and David Barrett, "Robust, Unconscious Self-Deception: Strategic and Flexible," *Philosophical Psychology*, vol. 29, no. 5, 2016, p. 682-96.

———, "Reply to Doody," *Philosophical Psychology*, vol. 30, no. 5, 2017, p. 677-681.

Gendler, Tamar, "Self-Deception as Pretense," in Gendler, Tamar, *Intuition, Imagination, and Philosophic Methodology*, Oxford, Oxford University Press, 2010a, p. 155-178.

———, "Alief and Belief," in Gendler, Tamar, *Intuition, Imagination, and Philosophic Methodology*, Oxford, Oxford University Press, 2010b, p. 255-281.

———, "Alief in Action and Reaction", *in* Gendler, Tamar, *Intuition, Imagination, and Philosophic Methodology*, Oxford, Oxford University Press, p. 282-310, 2010c.

Greco, Daniel, "A Puzzle about Epistemic Akrasia," *Philosophical Studies*, vol. 167, no. 2, 2014, p. 201-219.

Hieronymi, Pamela, "Controlling Attitudes," *Pacific Philosophical Quarterly*, vol. 87, no. 1, 2006, p. 45–74.

Holton, Richard, "What Is the Role of the Self in Self-Deception?" *Proceedings of the Aristotelian Society*, vol. 101, no. 1, 2001, p. 53-69.

Hookway, Christopher, "Epistemic *Akrasia* and Epistemic Virtue," in Fairweather, Abrol and Linda Zagzebski (eds.), *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*, Oxford, Oxford University Press, 2001, p. 178-99.

Hubbs, Graham, "Alief and Explanation," *Metaphilosophy*, vol. 44, no. 5, 2013, p. 604-620.

Johnston, Mark, "Self-Deception and the Nature of Mind," in McLaughlin, Brian and Amélie Oksenberg Rorty (eds.), *Perspectives on Self-Deception*, Berkeley, University of California Press, 1988, p. 63-191.

Lear, Jonathan, *Aristotle and the Desire to Understand*, Cambridge, Cambridge University Press, 1988.

Levy, Neil, "Self-Deception and Moral Responsibility," *Ratio*, vol. 17, no. 3, 2004, p. 294-311.

Lynch, Kevin, "On the 'Tension' Inherent in Self-Deception," *Philosophical Psychology*, vol. 25, no. 3, 2012, p. 433-450.

Mandelbaum, Eric, "Against Alief," *Philosophical Studies*, vol. 165, no. 1, 2013, p. 197–211.

Mele, Alfred, *Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control*, Oxford, Oxford University Press, 1987.

———, "Real Self-Deception," *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 91–102.

———, "Twisted Self-Deception," *Philosophical Psychology*, vol. 12, no. 2, 1999, p. 117–137.

———, *Self-Deception Unmasked*, Princeton, Princeton University Press, 2001.

Michel, Christoph, and Albert Newen, "Self-Deception as Pseudo-Rational Regulation of Belief," *Consciousness and Cognition*, vol. 19, no. 3, 2010, p. 731-44.

Nelkin, Dana, "Self-Deception, Motivation, and the Desire to Believe," *Pacific Philosophical Quarterly*, vol. 83, no. 4, 2002, p. 384-406.

Owens, David, "Epistemic Akrasia," *The Monist*, vol. 85, no. 3, 2002, p. 381-397.

Padel, Ruth, *In and Out of the Mind: Greek Images of the Tragic Self*, Princeton, Princeton University Press, 1992.

Patten, David, "How Do We Deceive Ourselves?" *Philosophical Psychology*, vol. 16, no. 2, 2003, p. 229-246.

Plato, *The Republic*, G.M.A. Grube (trans.), Indianapolis, Hackett, 1992.

Sanford, David, "Self-Deception as Rationalization," in McLaughlin, Brian and Amélie Oksenberg Rorty (eds.), *Perspectives on Self-Deception*, Berkeley, University of California Press, 1988, p. 157-170.

Scott-Kakures, Dion, "Self-Deception and Internal Irrationality," *Philosophy and Phenomenological Research*, vol. 56, no. 1, 1996, p. 31-56.

———, "Motivated Believing: Wishful and Unwelcome," *Nous*, vol. 34, no. 3, 2000, 348-375.

———, "At 'Permanent Risk': Reasoning and Self-Knowledge in Self-Deception," *Philosophy and Phenomenological Research*, vol. 65, no. 3, 2002, p. 576–603.

———, "Unsettling Questions: Cognitive Dissonance in Self-Deception," *Social Theory and Practice*, vol. 35, no. 1, 2009, p. 73-106.

Setiya, Kieran, "Believing at Will," *Midwest Studies in Philosophy*, vol. 32, 2008, p. 36-52.

Shah, Nishi, and J. David Velleman, "Doxastic Deliberation," *Philosophical Review*, vol. 114, no. 4, 2005, p. 497-534.

Soteriou, Matthew, "Mental Action and the Epistemology of Mind," *Nous*, vol. 39, no. 1, 2005, p. 83-105.

Soteriou, Matthew, and Lucy O'Brien (eds.), *Mental Action*, Oxford, Oxford University Press, 2009.

Talbott, William, "Intentional Self-Deception in a Single Coherent Self," *Philosophy and Phenomenological Research*, vol. 55, no. 1, 1995, p. 27–74.

Trope, Yaacov, and Nira Liberman, "Social Hypothesis Testing: Cognitive and Motivational Mechanisms," *in* Kruglanski, Arie. W. and E. Tory Higgins (eds.), *Social Psychology: Handbook of Basic Principles*, New York, Guilford, 1996, p. 72-101.

Williams, Bernard, "Deciding To Believe," in Williams, Bernard, *Problems of the Self*, New York, Cambridge University Press, 1981, p. 136-151.