teorema
Vol. XXVI/3, 2007, pp. 123-130
ISSN: 0210-1602

Cognitive Peers and Self-Deception[1]

Charles M. Hermes

RESUMEN

Alfred Mele ha revisado recientemente su test de pares cognitivos, que tuvo su origen en *Self Deception Unmasked*, para tratar un caso problemático. Mientras que la revisión es una versión mejorada del original, aún resulta susceptible de verse afectada por dos tipos de contraejemplos. Sin embargo, dado que el test captura algo esencial al autoengaño, deberíamos mejorar, en vez de abandonar el método de Mele. Por lo tanto, después de explorar los problemas que plantea su versión, sugeriré ciertas modificaciones que llevan a una versión más fuerte de la prueba.

ABSTRACT

Recently Alfred Mele revised his cognitive peers test, which originated from *Self Deception Unmasked,* to deal with a problematic case. While the revision is an improvement upon the original, it still falls prey to two types of counterexamples. Nevertheless, because the test captures something essential to self deception we should improve, instead of abandon, Mele's method. Therefore, after exploring problems with his account, I shall suggest modifications that lead to a stronger version of the cognitive peers test.

Recently Alfred Mele revised his cognitive peers test, which originated from *Self Deception Unmasked,* to deal with a problematic case. While the revision is an improvement upon the original, it still falls prey to two types of counterexamples. Nevertheless, because the test captures something essential to self-deception we should improve, instead of abandon, Mele's method. Therefore, after exploring problems with his account, I shall suggest modifications that lead to a stronger version of the cognitive peers test. Mele's original test claims:

> If S is self-deceived in believing that *p*, and D is the collection of relevant data readily available to S, then if D were made readily available to S's impartial cognitive peers (including merely hypothetical people), those who conclude that *p* is false would significantly outnumber those who believe that *p* is true [Mele (2001), p.106].

To be impartial, a cognitive peer must at least lack S's desire that p be true as well as lack any desire that ~p is true. Further a peer should not favor erring in either direction [Mele (2003), p. 107]. To be peers, they should have similar conceptual abilities. While cognitive defects could cause an agent to treat data in bizarre manners, relevant cognitive peers must share these problems. Finally, Mele has recently revised the test to incorporate the condition that peers spend at least as much time and effort evaluating D as the agent does [Mele (2001), p. 164].

Since, intuitively, self-deception is a species of motivated belief, the desires of the deceived agent must play a causal role. Someone who lacks the motivated aspects of the deceived would, therefore, lack an important factor in how the belief was derived. So, something in the spirit of Mele's test is necessary. As Mele notes, however, the test is open to at least one difficulty.

Suppose a scientist's wife is dying of a rare disease driving him to discover a cure with a zeal that surpasses his colleagues. While the same evidence might be available to the scientist and his cognitive peers, his peers might reach opposing conclusions because they lack the necessary motivation to extensively examine the data. While desires cause the scientist to carefully ponder data, they need not cloud his judgment. Clearly being motivated to examine data should not be sufficient for being motivationally biased in the pejorative sense pertinent for self deception. By stipulating that cognitive peers must spend at least as much time and effort examining evidence, this problem is avoided[2].

Unfortunately, a more serious problem occurs in cases of overdetermination. Suppose John's late nights at work cause his wife, Sue, to worry that he is having an affair with a coworker. While her husband is supposedly at work, Sue sees John out for a drive with a female passenger and interprets this as strong evidence of John's fidelity (after all, the car is not at a motel room). The following day, when John has not returned home late into the night, Sue drives to his work place. Seeing only John's car and the car of his attractive secretary in the parking lot, Sue becomes further assured that John is merely working. That weekend John presents Sue with a new car, explains that the woman she saw was the dealer, and apologizes for the late nights working to pay for the present. By the weekend, both Sue and her cognitive peers form the false belief that John is faithful.

Still, Sue treats data in a drastically different manner than her cognitive peers. Most peers' faith in John would diminish in the first two days, yet Sue takes these as evidence of his innocence. Since the only significant distinction between Sue and her cognitive peers is their desires, desire plays a significant role in Sue's belief formation. This suggests that Sue is motivationally biased.

To account for this case the test must consider not only the final decision, but also degrees of confidence. While Sue derived the same belief as her

cognitive peers, slight modifications to the story would have resulted in a different conclusion. If the last piece of evidence were less convincing, Sue, unlike her cognitive peers, would have been persuaded. In order to account for overdetermination cases, not only must the agent pass the original test, but changes in the relevant data cannot alter agreement with cognitive peers. If an agent could have the same evidence as her cognitive peers and derive a different conclusion, something must be causing the discrepancy.

Since the only relevant difference between an agent and her cognitive peers is their desires, discrepancies could only be caused by motivational biases. Therefore, the test could be strengthened by allowing p-relevant data to be added or subtracted (or both) from D. In Sue's case, replacing the third bit of evidence with something weaker would have resulted in a discrepancy between her and her cognitive peers. So, this revision correctly characterizes Sue as motivationally biased.

While the above modification solves Sue's case, more problems abound for Mele's test. As Mele aptly notes; "Our desiring that *p* can contribute in a variety of ways to our believing that *p* in instances of strict self deception" [Mele (2001), p. 25]. Unfortunately the revised test, as well as the original, fails to capture some of the ways that desires bias beliefs. While the revised test adequately accounts for situations in which a desire alters the assessment of evidence, it fails to capture all cases of motivationally biased evidence gathering.

A problematic case of this sort can be derived from a slight modification of one of Mele's paradigm cases of self-deception. The case depicts Beth, who spends more time viewing photos of her playing with her father than photos of her brothers doing likewise [Mele (2001), p. 18]. She also often reflects upon fond memories of her and her father, thereby making those memories more vivid. Unlike Mele's case, suppose Beth would never contemplate whether her father loved her most; unless prompted. Nonetheless, years later, after the photo album is lost, her brother asks her if their father loved her most. Desiring to be truthful forces Beth to be objective about all the evidence she can gather.

Unfortunately since all tangible evidence is gone, she must rely solely upon her memories. Seeing that the most abundant and vivid of these depict photographs of her and her father she forms the belief that her father loved her most. Since her assessment of the data is objective, her cognitive peers would derive the same conclusion. Still, it appears that motivational bias plays a significant role in how Beth derives her belief. Yet, since she forms the same conclusion as her cognitive peers would form upon all the available data, Mele's cognitive peers test indicates that she is not properly motivationally biased for being susceptible to self-deception.

In order to avoid this difficulty, it is essential to notice that the effect a desire has on the formation of a belief can occur in a pre-doxastic state. A de-

sire can influence not only the assessment of gathered data but also the process of collecting data. In order for a test to correctly capture the revised Beth case it must be sensitive to all periods in which desire could play a causal role in belief formation.

To incorporate this sensitivity, the relevant data that an agent shares with her cognitive peers needs to be extended. Instead of merely providing cognitive peers with the set of data that is currently available to the agent when she forms the belief, the p-relevant data that the agent has throughout the deliberation process must be made available to her cognitive peers at those times.

Suppose Beth's cognitive peers had access to the photo album whenever Beth did. The manner in which they would view the album would be quite different than Beth's. Without the added pleasure of viewing pictures of Beth playing with her father, they lack reason to linger on these photos. Therefore, their memory of the album would be distinct from Beth's. Unfortunately, it might appear difficult to discern how her cognitive peers would examine the photo album. If someone is presented with a strangers photo album the manner in which they would peruse it would be quickly — if at all. If Beth's cognitive peers only see the album through her eyes, then they too would be focusing upon pictures of her and her father. If their motive is limited to determining whether Beth's father loved her most, they would catch factors that a normal agent would miss; thereby making the test trivially passable by all normal agents.

Nevertheless, it is not necessary that Beth's cognitive peers lack all of her desires. In fact, being a cognitive peer implies having the same desires except for those directly relevant to determining whether Beth's father loved her most. Therefore, there usually would be ample reason for cognitive peers to pay some attention to evidence. In the remaining cases, where the only salient reason for observing a piece of evidence is due to a desire to form a certain belief, the test's result of deeming the belief motivationally biased seems appropriate.

Even if Beth's case did not force the test to be sensitive to the history of evidence gathering there are other reasons to build in historical sensitivity. Perhaps the phenomenon of belief persistence is always accompanied by a desire to maintain the belief being defended. Nonetheless, whether or not this is the case should not be determined *a priori*. If a test produced for determining when someone is motivationally biased entails a position on this issue, then the test proves too much. As I shall show, without historical sensitivity, the cognitive peer test suggests that belief persistence is always accompanied by a desire[3].

Suppose students normally take introduction to philosophy classes before taking nineteenth century philosophy and instructors of nineteenth century philosophy are typically charitable to Hegel while instructors of introduction to

philosophy are hostile. Further suppose Joe, ignoring his academic advisor, took nineteenth century prior to introduction to philosophy. Unlike other students who enrolled in both courses, Joe maintains that Hegel is one of the greatest philosophers. Yet, Joe has all the same information and a very similar cognitive make up as many of his classmates. In fact, the sole reason for the difference between his belief and his classmate's is the order in which they received information. Still, Joe does not appear to be more susceptible to motivational bias than his classmates.

If belief persistence is always accompanied by a desire that the persisting belief is true, then a test that is not historically sensitive would be passed by both Joe and his classmates. Since cognitive peers, lacking the relevant desire, would not experience belief persistence and the evidence on both sides is equal (pre hypothesis), they would remain neutral. Therefore, both Joe and his classmates would derive different decisions than their impartial cognitive peers.

If belief persistence is not always accompanied by a desire, Joe's classmates count as his cognitive peers. Since they have all the same data and have no p-relevant desire they meet the conditions of an impartial cognitive peer. Since the majority of Joe's cognitive peers (his classmates) disagree with him, his belief is motivationally biased while theirs is not. Yet, clearly Joe is not in a worse epistemic position than his classmates. So, if belief persistence is not always accompanied by a desire we achieve absurd results.

The problem occurs because the test is not sensitive to the history of data collection. If Joe's classmates received all the same data at the same times they would have formed the same belief as Joe. Assuming that belief persistence is purely cognitive, Joe's cognitive peers would also derive the same belief, since they would be susceptible to belief persistence. In this case, both Joe and his classmates would not be motivationally biased. If belief persistence is desire based, then cognitive peers would not experience belief persistence. In this case the test would deem that both Joe and his classmates were motivational biased.

Whether or not belief persistence is always accompanied by a desire is a matter that should not be settled merely by a test for motivationally biased beliefs. Nonetheless, unless the test is historically sensitive, it entails that belief persistence is always accompanied by a relevant desire. Therefore, not only is a test that is not historically sensitive too weak to properly capture Beth's case, but, by making *a priori* judgments where *a posteriori* work is clearly required, it is also too strong.

Still there may be a concern that if belief persistence is ubiquitous it trivializes the test. If a desire is always involved in belief persistence then the result that any agent experiencing belief persistence has a motivationally biased belief seems right. After all, their desire is playing a causal role in the maintenance of their belief. Unfortunately, since the test is intended as a

method for determining whether an agent possesses a necessary condition for self deception this result seems problematic. Clearly the test needs to be capable of ruling out some candidates. Here, appeal must be made to degrees. If there is a standard effect that desire plays in all beliefs, then when we attribute a biased belief to someone in a disapproving manner we are attributing to them a belief that is influenced by desire to a degree greater than normal.

Another problem for the revised test seems to occur because an agent who originally is self-deceived in deriving a belief later may obtain stronger evidence. After gaining new data, it might be doubted that the agent is still motivationally biased. Nonetheless, the revised test removes the possibility that mere addition of evidence could eliminate a motivational bias. An agent who once passed the test for motivational bias will continue doing so regardless of any new information that is gathered. There will always remain a set of evidence that if they and their cognitive peers shared would yield different conclusions about p, namely the set of evidence that caused them to originally pass the test. This would be problematic if motivational bias were sufficient for self deception. Since it is only a necessary condition for self deception, an account of self deception could accept that these individuals are no longer self-deceived while maintaining that they are still motivationally biased.

This seems to be the correct result. Gaining new information does not reduce the motivational bias that the agent originally had. It may, however, reduce how important the causal role of the motivational bias is upon the formation of the belief. Further evidence supporting the conclusion that was originally derived from motivational bias merely overdetermines the belief. Nonetheless, as long as part of the data supporting the belief is affected by a relevant desire, the belief still has a motivationally biased aspect. In order to eliminate ones motivational bias, more needs to occur than simply adding new information. The way information is treated must also change.

One may still be concerned that while the revised test is necessary for motivational bias, it does not appear sufficient. If someone would behave identically in all hypothetical cases toward a belief, regardless of whether they posses a relevant desire, then desire is not playing a causal role. Therefore, passing the revised test is clearly a necessary condition for having a motivationally biased belief.

Suppose, however, that certain evidence is highly sensitive to the effects of a desire while other evidence is not. Perhaps some pieces of information require more interpretation than others making the influence of desire more effective. An agent then could have only information that is highly immune to the effects of desires when forming a belief. Possessing only this kind of information one could form a belief that p without desire playing a causal role. Nonetheless, there could be another piece of evidence, not possessed by the agent, which she would have weighed improperly because of a p relevant desire. Such a person would pass the revised test because if she ob-

tained that piece of evidence her belief in p would diverge from her impartial cognitive peers. Nevertheless, since the desire is not playing an active role in the formation of the belief we might not count the individual as biased.

Still, in these cases, the agent would act radically different under certain counterfactual circumstances because of her desire. Further, motivational bias appears to be a dispositional property. Motivational bias usually does not manifest itself at all moments the bias is present. Instead, someone with a motivational bias is disposed to act in certain manners when other conditions obtain. This dispositional character of motivational bias makes counterfactual circumstances in which an agent is disposed to treat evidence inappropriately, but not currently doing so, pertinent. This leads to two different kinds of motivational bias. The bias could be either active or dormant. Since it seems that all cases of self-deception require active motivational bias, this points toward another necessary condition for self-deception. Nevertheless, since the goal of this paper is to improve upon a test for determining whether an agent is motivationally biased, the task of further differentiating active and dormant motivational bias falls outside its scope. The above considerations demonstrate that dormant motivational bias should be counted as a species of motivational bias. Therefore, it is a virtue of the revised test that it includes these cases.

While Mele's impartial peers test captures an important aspect of self deception, revision is necessary to avoid problematic cases. Through extending the sets of evidence that are pertinent for determining motivational bias and becoming sensitive to the history of evidence collection a more accurate test for motivational bias can be derived.

Department of Philosophy
University of Texas at Arlington
Arlington, TX 76019, USA
E-mail: hermes@uta.edu

NOTES

[1] I would like to thank Alfred Mele and Jason Turner for extensive comments on an earlier draft of this paper.

[2] This case and solution parallel Mele's [Mele (2001), pp. 164-165].

[3] Mele claims that the cognitive peer's test is a test for determining whether someone is appropriately motivationally biased for being self-deception. The following argument treats the test as a sufficient condition for being appropriately motivationally biased. Perhaps Mele would respond that the test should not be treated as a sufficient condition for motivational bias, but as something weaker. Nevertheless, once historical sensitivity is built into the test, it appropriately accommodates more cases.

REFERENCES

HATZIMOYSIS, A. (ed.) (2003), *Philosophy and the Emotions*: Cambridge University Press.
MELE, A. (2001), *Self Deception Unmasked:* Princeton University Press.
— (2003), "Emotion and Desire in Self-Deception", in Hatzimoysis (ed.) (2003), pp.163-179.