

MIT OpenCourseWare
<http://ocw.mit.edu>

24.120 Moral Psychology
Spring 2009

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

XXI Self-Deception

EMPIRICAL FINDINGS ON SELF-DECEPTION

A survey of one million high school seniors found that 70% thought they were above average in leadership ability, and only 2% thought they were below average. In terms of ability to along with others, all students thought they were above average, 60% thought they were in the top 10%, and 25% thought they were in the top 1% ... A survey of university professors found that 94% thought they were better at their jobs than their average colleague.

Gilovich, reported in Mele

In a study by Kunda, women who were heavy consumers of caffeine were less likely to believe a study that said that caffeine increased the risk of breast cancer than were low consuming women or men.

An old explanation

If a man thinketh himself to be something, when he is nothing, then he deceiveth himself (Gal. 6: 3-4)

DECEPTION BY OTHERS

What is involved in a standard case of deception? A deceives B about m, just in case A knows some proposition p about m, yet A intentionally brings B to believe that not-p.

SELF-DECEPTION MODELLED ON THE DECEPTION OF OTHERS

A deceives herself about m, just in case A knows some proposition p about m, yet A intentionally brings herself to believe that not-p.

Worries: (i) how can A know that p, and yet believe that not-p?
 (ii) how can A intentionally manipulate her own beliefs?

In answer to the first worry, perhaps one could forget one's initial belief. Mele's example of Ike, the diary writer. But if this is to be effective self-deception, typically evidence to the contrary must be constantly diverted from one's attention; and that means that one must retain knowledge of the kind of evidence that must be covered up.

THE DIVIDED SELF

Freud: at its core Freud's is a theory of conscious self-ignorance: one is not aware of many of one's drives and desires since they are in the unconscious. However, it also contains a mechanism of self-deception. In Freud's later theory this comes out in the tripartite division of the self into the id, super-ego and ego. The id is the unconscious source of our drives. The ego is the conscious self. Many of the drives of the id are too disturbing for the ego to cope with. So the super-ego comes to repress these over-disturbing drives. (Or, since the super-ego also contains

the conscience, it represses them simply because it doesn't approve of them.) Repressed drives are kept back in the unconscious.

What should we identify the agent with? How agent like is the super-ego? Johnston's example of the self-deceptive belief that one is still sober enough to drive the car.

MELE

An alternative is to reject the idea that self-deception involves an intention on the part of the agent to lead herself astray. Maybe it is just that the agent is deceived, i.e. mistaken, in some particular way.

Mele's four jointly sufficient conditions for *entering self-deception in acquiring a belief that p*:

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

Mele's putative four sources of motivated bias:

Negative misinterpretation
Positive misinterpretation
Selective focusing
Selective evidence gathering

Three more well documented sources of bias:

vividness of information
availability heuristic
confirmation bias

DOES THE BIAS NEED TO BE MOTIVATED?

A first pass: A 's belief in p is motivated just in case either A wants p to be true, or A wants to believe that p .

Problem: the under-self-confident agent.

Perhaps such an agent wants to believe that she is no good in order to avoid having to face failure. But it seems unwarranted to require that this is always the case.

An alternative: the belief is motivated just in the sense that maintaining it caused minimum disruption to the agent's belief system.

ARE MELE'S CONDITIONS SUFFICIENT FOR SELF DECEPTION?

Consider the case of Jean-Marie:

Jean-Marie is a racist. He thinks that blacks and Arabs are not as good as whites: not as clever, or as imaginative, or as brave, or as trustworthy, or whatever. Take just about any property that Jean-Marie might regard as a virtue, and he will think that whites have more of it. Let us assume that his beliefs here are, by and large, false. But he holds them sincerely. And were we to challenge them, he would provide evidence: reams of it, taken from the magazines and newspapers of the kinds of organization to which he belongs. He is aware of the opposing view; indeed he has reams of that too, collected to document the conspiracy which he thinks pervades the liberal establishment that controls the mainstream press and publishing houses.

Jean-Marie meets all four of Mele's conditions. He is bigoted and prejudiced. Yet he is not obviously self-deceived. Perhaps the reason is that he lacks false beliefs about himself. Or does he? The more we reflect on his likely self-ignorance, and its causes, the more likely we are to say that he does have false beliefs about the reliability of his beliefs; but then the more likely we are to say that he is self-deceived. So consider the case of Catherine:

Catherine has applied for several jobs recently, and has been unsuccessful each time. She has also been horribly disappointed each time. She puts her disappointment down to too much thought. On each occasion she had spent a great deal of time thinking about the job, had, as a result, imagined just what it would be like to get it, and so had been devastated when she didn't. She has just decided to apply for another job. She thinks that it is clearly better than her current job; otherwise she wouldn't be putting in for it. But she has resolved not to think too deeply about what it is like, or to examine the evidence that she has; at least not until or unless she gets an offer. She knows that, were she to think more about the job, there is some chance that her opinion of its merits would change; but she thinks that the possibility of error here is worth risking to maintain her equanimity. As it happens she is radically wrong about the job. It is a terrible job, far worse than her current one, as a little more reflection would have shown her.

A proposed amendment to Mele's conditions:

S is self-deceived about a subject matter α if

1. *S*'s body of beliefs about α contains mistaken beliefs about the self.
2. *S* treats data relevant, or at least seemingly relevant, to the truth value of these mistaken beliefs in a motivationally biased way.
3. This biased treatment is a nondeviant cause of *S*'s acquiring these mistaken beliefs.
4. The body of data available to *S* at the time provides greater warrant for rejecting these beliefs than for accepting them