

PHIL 5983: Rationality Seminar

University of Arkansas, Fall 2004

Topic: Self-Deception I: The Puzzles of Self-Deception

Readings: *Self-Deception Unmasked* (Chapters 1 and 2); Davidson's "Deception and Division"

*Bryan will cover the Davidson reading.

Chapter 1

*Self-deception involves some kind of motivated and biased belief manipulation. However, Mele denies that *self*-deception should be understood on the model of *interpersonal* deception.

--First, Mele distinguishes between the following two kinds of motivation for self-deception:

straight self-deception: Self-deception in which the agent deceives herself about something that she wants to be the case. (Typical)

twisted self-deception: Self-deception in which the agent deceives herself about something that she does not want to be the case. (Atypical)

--The category of self-deception can be captured: 1) lexically, 2) by examples, or 3) by theory (or some combination thereof).

--Beginning with a definition of 'deceive', Mele argues that we can be misled into thinking of self-deception on the interpersonal model. And Mele argues that this leads to 2 puzzles.

Here are two assumptions about the meaning of 'deceive':

1. By definition, person A deceives person B (where B may or may not be the same person as A) into believing that p only if A knows, or at least believes truly, that $\sim p$ and causes B to believe that p.
2. By definition, deceiving is an intentional activity: nonintentional deceiving is conceptually impossible." (6)

These two assumptions generate the following 2 puzzles of self-deception:

static puzzle: How can a self-deceiver both believe that p and believe that not-p? (Or even worse, how can a self-deceiver believe that p and not believe that p?!)

dynamic puzzle: How can a self-deceiver intentionally deceive himself? How can the self-deceived simultaneously possess the sophistication to trick oneself, and the ineptitude to fall for these tricks?

Mele will avoid these puzzles by denying the assumptions.

*Let's say some more about motivationally biased belief manipulation. This manipulation can occur at various levels: evidence gathering, hypothesis formation, inferences, or belief retention (compare this with the bottom of p. 11).

--Not all motivated irrationality (bias) is intentional, however. Mele presents the experimental study about women caffeine drinkers' reactions to a reported link between heavy caffeine intake and disease in women in order to illustrate this point. (12-13)

Some motivated irrationality is intentional, some is not. Contrast this point with the agency/antiagency dichotomy presented on p. 13. One might try to salvage the agency view, for example in accounting for self-deception, by appealing to unconscious intentions. Mele doesn't deny the possibility of such intentions, but does not think they are likely in self-deception. He recognizes that he needs to give an account of how the desire motivates without an intention.

--Note the 3 kinds of motivation that Mele distinguishes on the top of p. 18.

Q: Are the cases on p. 20 (e.g., Bob and the teenager) really ones in which there is no motive to *believe*?

Chapter 2

*So, how can a desire that p lead to a belief that p? Mele provides 4 examples: negative misinterpretation, positive misinterpretation, selective focusing/attending, and selective evidence gathering. (26-27)

--Notice that these examples do not require that the person originally had a contradictory belief. And Mele calls these cases of self-deception.

--We can distinguish "hot" (motivated) and "cold" (unmotivated) biases. Tversky and Kahneman introduced us to several cold biases, and Mele discusses 3 of these on pp. 28-29. Mele also notes that these cold biases can be incorporated into a hot bias project as well.

*Mele introduces the PEDMIN model of hypothesis testing as the central piece to his account of self-deception. According to this picture, we generate and test hypotheses (and eventually come to believe) with the goal of avoiding costly errors. According to this model, it is more important that we avoid falsely believing something that will harm us or cause us discomfort, than it is to believe the truth.

--Here is the quick application to self-deception: It would be a costly error mistakenly to believe that you are deficient in some way (e.g., below average in

your profession) or to believe something that is discomfoting (e.g., that your spouse is having an affair).

--You can always get things wrong in 2 different ways—mistakenly believing that p or mistakenly believing that not-p. However, one way of getting things wrong often “hurts” more, and for this reason we will often have different thresholds of belief for these 2 possibilities. These thresholds, naturally, are relative and vary from person to person—recall the treason case, pp. 36-37.

Q: Is PEDMIN the general belief-formation rule, or does good, old-fashioned respect for truth sometimes rule?

--Mele points out that his endorsement of PEDMIN is consistent with his antiagency accounts of motivated biases.