Book Review

[(2013) *Cognitive Neuropsychiatry* 18(1-2): 146-151; please cite published version]

Robert Trivers (2011). *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*. New York: Basic Books. Pp. 397. ISBN 978-0-4650275502. $28.00 US.

It's an intriguing evolutionary puzzle. Why, we may ask, would natural selection equip us with sensory and reasoning capacities designed to give accurate information about the world, only to give us a trait—the capacity for self-deception—that undermines that accurate information at critical moments in our lives? It's as if evolution were a cruel parent, who gives us sophisticated cognitive toys but restricts our ability to enjoy them properly. And the fitness value of accurate information *seems* so obvious: survival and reproduction depend on knowing where to find food, where the predators are, who your mates are, who your enemy is, and what your abilities are for dealing with the challenges of your environment. But self-deception seems to threaten such hard-won informational gains. Why was it not selected out?

Anyone who feels the pull of this puzzle will be delighted by the appearance of *The Folly of Fools: The Logic of Deceit and Self-Deception in Human Life*, written by legendary evolutionary theorist Robert Trivers. Trivers himself puts the puzzle this way:

> Together our sensory systems are organized to give us a detailed and accurate view of reality, exactly as we would expect if truth about the outside world helps us navigate it more effectively. But once this information arrives in our brains, it is often distorted and biased to our conscious minds. We deny the truth to ourselves. We project onto others traits that are in fact true of ourselves—and then attack them! We repress painful memories, create completely false ones, rationalize immoral behavior, act repeatedly to boost positive self-opinion, and show a suite of ego-defense mechanisms. Why? (p. 2)

It's worth noting further that your typical commonsense explanation for the existence of self-deception—that it helps keep us happy by shielding us from painful truth—doesn't stand a chance at solving the evolutionary puzzle. And that's for a simple reason: natural selection doesn't care about your happiness; it only cares about survival and reproduction.

What then is the solution? If we follow Trivers, we'll first note that there can be evolutionary advantages to deceiving *others*. Appearing stronger than we are may scare off a rival. Appearing saintly, despite moral turpitude, may incur favors from conspecifics. Appearing more in love than we really are may charm a mate. Second, we'll note that conscious lies may be hard to maintain—lies can tell on the face and in word choice. Third, we'll note that the believer of a falsehood may propagate it more convincingly than a disbeliever. And now, with all these pieces in place, we get Trivers' enticing proposal: the evolutionary origins of the human

propensity for self-deception lie in the adaptive benefits of deceiving others. Or, "we deceive ourselves the better to deceive others" (p. 3).

In my view, this theory is both elegant and important. Furthermore, I suspect it may even be true of *some* of the phenomena that fall loosely under that paradoxical-seeming term "self-deception."

But it is the very promise of this theory that sets the reader up for disappointment. *The Folly of Fools* suffers from two remarkable flaws.

1. The book as a whole simply does not add up to being an *argument* for its main thesis.

2. The focal term "self-deception" is used so loosely throughout that it becomes impossible to determine what the scope of the thesis actually is.

The second flaw of course contributes to the first. So let's start with the second.

We'll ask a simple question. Is an overconfident football player really in the grip of the *same* psychological phenomenon as a gambling addict who tells himself he doesn't have a problem? Maybe, but maybe not. Of course we might—loosely—label both of them cases of "self-deception." But that may be like calling Venus a "star," using the same appellation for a large satellite of the sun as we do for larger burning balls of gas light years away. Same pre-scientific word—"star"—but distinct phenomena. It was of course an achievement in astronomy to realize that Venus was not the same sort of thing as other stars. And this points to a major desideratum on scientific enterprises: *recognizing genuinely distinct phenomena, despite the conflations of everyday speech*.

Trivers, however, appears unconcerned with this desideratum. He does, to be fair, attempt a definition of self-deception: "true information is preferentially excluded from consciousness . . . false information is put into the conscious mind" (p. 9). But this is vague, and he goes on in the course of the book to lump so many things under the heading "self-deception" that every bias psychology has ever discovered seems to count as self-deception. Examples are in order. On page 14, Trivers lists primate "implicit in-group favoritism" among "monkey forms of self-deception." On page 65, he brings up the well-known phenomenon of stereotype threat (members of a stereotyped minority perform worse on a range of tasks when reminded of their minority status) and refers to it as an "imposed self-deception." On pages 70-71 he refers to patients' susceptibility to the placebo effect as "self-beneficial self-deception." On page 100, he describes behaviors that "may also involve self-deception" and includes those stemming from "false emotion," by which he means feelings of romantic love in a male prior to sex that dissipate soon after. On page 104 female "temporary fantasy" at the time of ovulation about other sexual partners gets counted as "a voluntary, conscious kind of self-deception." On page 159, male overconfidence is counted as self-deception. On page 177, falling for the flattery of a con man, like Bernie Madoff, is attributed to self-deception. Chapter 10 is dedicated to societal "false historical narratives," which are counted as self-deception. And Chapters 11 and 12 are dedicated to self-deception in going to war and self-deception in religion, respectively.

This is all more heat than light. I'm not claiming that any phenomena mentioned in the previous paragraph are *wrongly* termed "self-deception." Rather, I'm claiming that if the category of *self-deception* is so diffuse as to include them *all*, that category becomes next to useless for purposes of scientific theory and investigation.[1] Alternately: if we don't have evidence that there are common neural or psychological mechanisms implicated in all these diverse phenomena, we don't have much reason to treat them as a unified class.

Furthermore, this lack of definitional clarity can muddle otherwise important questions about phenomena to which we might consider extending Trivers' theory. A reader of this journal, for example, might wonder to what extent schizophrenia or various monothematic delusions, like Capgras', are to be counted as "self-deceptive." Alternately, one might wonder whether the feelings of grandeur that occur at high points in the bipolar cycle should be thought of as "self-deception," or whether one should put any of the anosognosias following neurological damage in this category. Or, is body dysmorphia a form of self-deception? We could raise analogous questions about a host of other clinical phenomena ("Is X a form of self-deception?"). From a book about deceit and self-deception, we might hope for some psychological theory that gives clear guidance on such questions. But that hope would be vain in the present case.

Now we come to the first major flaw: *The Folly of Fools* does not provide an argument for its main thesis. What would a proper argument look like? Since Trivers can't rely on "self-deception" as a unified class, for each type of "self-deception" he would like to include under his main thesis he would have to do the following:

1) Develop *specific* hypotheses about the evolutionary benefits and costs of a particular Self-Deceptive Phenotype as opposed to the benefits and costs of a Liar Phenotype *and* as opposed to an Honest Phenotype within the same ecological niche.

2) Show, through psychological research, that a large portion of humans has the Self-Deceptive Phenotype so posited.

3) Find empirical evidence for the hypotheses' entailments about costs and benefits.[2]

1 is a requirement on having a theory that appeals to natural selection at all; 2 is needed for that theory to apply to real humans; and 3 is needed to support the claim that a given phenotype was selected *for* something, in this case deception[3].

---

[1] Steven Pinker (2011) voices a similar concern, referring to the concept of self-deception as "apparently profligate."

[2] This list is partially inspired by Lloyd (2001).

[3] A competing proposal is that self-deception is a byproduct of other traits with no function of its own (Van Leeuwen 2007; Mercier 2011).

Let's consider, for example, male self-presentation in courtship, where Trivers thinks overconfidence is a form of self-deception evolved in part to attract females. This hypothesis yields some testable predictions in keeping with the above requirements, a few of which are:

> **Prediction 1**: the presence of desirable females triggers self-inflating self-deception in males.
>
> **Prediction 2**: males who self-deceptively self-inflate are preferred by females to simply honest males or males who knowingly lie.
>
> **Prediction 3**: males simply lying about their qualities will be found out more frequently than males self-deceived about their qualities.
>
> **Prediction 4**: retaliation against self-deceived overconfident males is typically less severe than against males who are discovered to have lied.

For all we know, these four predictions, or some of them, could turn out true. But we won't learn whether they're true from reading Trivers, who relies largely on anecdotes. For example:

> I am walking down the street with a younger, attractive woman, trying to amuse her enough that she will permit me to remain nearby. Then I see an old man on the other side of her, white hair, ugly, face falling apart, walking poorly, indeed shambling, yet keeping perfect pace with us—he is, in fact, my reflection in the store windows we are passing. Real me is seen as ugly by self-deceived me. (pp. 17-18)

This is an admirable instance of personal honesty—and the book is full of them—but the reader finishes the book wondering to what extent such anecdotes can be backed by more rigorous research.

Someone reading this book review so far might have the impression that Trivers neglects empirical data altogether. That's not my claim; the book in fact discusses quite a range of interesting research. My claim is rather that evidence that would support Trivers' specific theory is missing, over and above the relatively uncontroversial claim that self-deception exists. David Dunning (2011: 18) makes much the same point in his commentary on the recent *BBS* article that William von Hippel and Trivers (2011) co-authored advocating the same thesis as Trivers' book:

> The hypothesis, however, lacks one characteristic I wish it had more of—data. That is, the hypothesis is not completely new, having been forwarded, in some form or another, over that last quarter century (Trivers 1985; 1991), and so it could profit now from direct data that potentially support it rather than from additional weaving of indirect arguments and findings such as those the authors have spun here. . . . it should be easy to create experiments to see if people are more persuasive to others to the extent they have persuaded themselves of some untruth first.

So *The Folly of Fools* has two big flaws: lack of conceptual clarity and lack of an argument that would connect Trivers' theory to the evidence it needs. I wish to conclude, however, by emphasizing that these flaws do not entail that the book is devoid of interest, nor do they entail that its main thesis (suitably refined) won't turn out true. Among the most interesting features of the book are its descriptions of deception in non-human animals (Chapter 2), its discussion of genomic imprinting and parent-offspring conflict (Chapter 4), its discussions of the relations between psychology and immunology (Chapter 6), and its analyses of the self-deceptions at the root of airplane and space shuttle disasters (Chapter 9), all of which provide useful gateways into the relevant empirical literatures. Nevertheless, the argument I was hoping to find didn't greet me. I began this book review with the metaphor that evolution is a cruel parent, who gives us sophisticated cognitive toys but doesn't let us enjoy them properly. At the end of *The Folly of Fools*, our parent's cruelty is still unexplained.

Neil Van Leeuwen
Georgia State University
University of Johannesburg

Acknowledgement
I'd like to thank Ryan McKay for feedback and encouragement on this book review.

References

Dunning, D. (2011) "Get thee to a laboratory" *Commentary* on target article "The evolution and psychology of self-deception" by William von Hippel and Robert Trivers, *Behavioral and Brain Sciences* 34(1): 18-19.

Lloyd, E. (2001) "Science Gone Astray: Evolution and Rape" (review of R. Thornhill and C. Palmer *A Natural History of Rape*, Cambridge: MIT Press), *Michigan Law Review* 99(6): 1536-1559.

Mercier, H. (2011) "Self-deception: Adaptation or by-product?" *Commentary* on target article "The evolution and psychology of self-deception" by William von Hippel and Robert Trivers, *Behavioral and Brain Sciences* 34(1): 35.

Pinker, S. (2011) "Representations and decision rules in the theory of self-deception," *Commentary* on target article "The evolution and psychology of self-deception" by William von Hippel and Robert Trivers, *Behavioral and Brain Sciences* 34(1): 35-37.

Trivers, R. (1985) "Deceit and self-deception," in *Social Evolution*, Benjamin/Cummings: 395-420.

Trivers R. (1991) "Deceit and self-deception: The relationship between communication and consciousness," in *Man and beast revisited*, ed. M. Robinson and T. L. Tiger, Smithsonian Press: 175-91.

Van Leeuwen, D. S. N. (2007) "The Spandrels of Self-Deception: Prospects for a Biological Theory of a Mental Phenomenon," *Philosophical Psychology* 20(3): 329-348.

von Hippel, W. and Trivers, R. (2011) "The evolution and psychology of self-deception" (with *Commentary*), *Behavioral and Brain Sciences* 34(1): 1-56.