Cognition 115 (2010) 268–281

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Self-deception requires vagueness

Steven A. Sloman [a,*], Philip M. Fernbach [a], York Hagmayer [b]

[a] Department of Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI, USA
[b] Department of Psychology, University of Göttingen, Göttingen, Germany

ARTICLE INFO

Article history:
Received 1 June 2009
Revised 19 November 2009
Accepted 22 December 2009

Keywords:
Self-deception
Causal models
Causal decision theory
Diagnostic reasoning
Intervention

ABSTRACT

The paper sets out to reveal conditions enabling diagnostic self-deception, people's tendency to deceive themselves about the diagnostic value of their own actions. We characterize different types of self-deception in terms of the distinction between intervention and observation in causal reasoning. One type arises when people intervene but choose to view their actions as observations in order to find support for a self-serving diagnosis. We hypothesized that such self-deception depends on imprecision in the environment that allows leeway to represent one's own actions as either observations or interventions. Four experiments tested this idea using a dot-tracking task. Participants were told to go as quickly as they could and that going fast indicated either above-average or below-average intelligence. Precision was manipulated by varying the vagueness in feedback about performance. As predicted, self-deception was observed only when feedback on the task used vague terms rather than precise values. The diagnosticity of the feedback did not matter.

1. Introduction

Any psychotherapist will tell you that self-deception is common, indeed quite normal. It comes in a variety of forms. Our focus is limited to self-serving attributions (cf. Miller & Ross, 1975; Weiner, 1992), the human tendency to deceive oneself about the diagnostic value of one's own actions. The evidence for diagnostic self-deception of this kind remains mostly anecdotal (Baumeister, 1993; Goleman, 1985). We suspect that self-deception is not easy to obtain in the laboratory because it requires the subject to finely balance a set of mutually contradictory actions and beliefs. Subjects must lie to themselves successfully (Paulhus, 2008); that is, they must remain unaware that they are treating information that they know to be diagnostic as non-diagnostic or vice versa.

The analysis of self-deception is tricky not only because a falsehood is at stake, but because it concerns an agent's action. We can represent our own actions to ourselves in different ways. An action can be represented as either an intervention or an observation (Sloman & Hagmayer, 2006; see Fig. 1). To represent an action as an intervention involves treating it as a deliberate choice. For instance, an addict who treats her drug-taking as an intervention believes that she can stop at any time. In contrast, treating one's own action as an observation involves taking the same perspective on it that an outside observer has (cf. Bem, 1967), seeing the action as the result of external and internal forces impinging on the individual. In the case of addiction, such factors might include the person's addiction and the availability of the drug. The observational stance reduces the responsibility of the actor because it attributes the action to forces other than the actor's free will.

In this paper, inspired by the plethora of evidence that reasoning can be biased by motivation (e.g., Dawson, Gilovich, & Regan, 2002; Dunning, 1999; Kunda, 1990), we propose that it is this ambiguity in how people represent their own actions that allows self-deception to arise. Actions can be construed as an agent's willful intervention or as the effect of factors governing the agent. The

* Corresponding author. Address: Cognitive and Linguistic Sciences, Brown University, Box 1978, Providence, RI 02912, USA. Tel.: +1 401 863 7595; fax: +1 401 863 2255.
E-mail address: steven_sloman@brown.edu (S.A. Sloman).

0010-0277/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2009.12.017

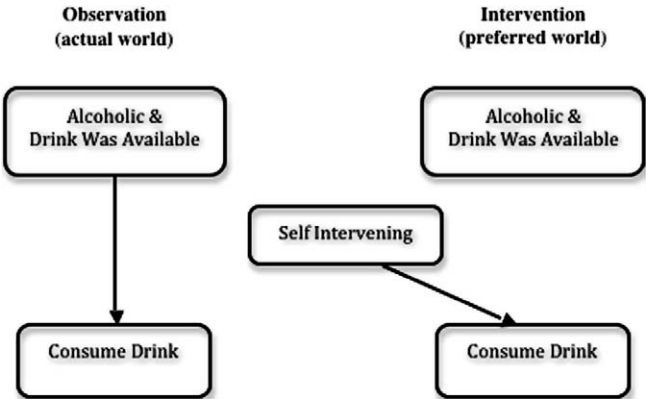

Fig. 1. Self-deception in addiction.

existence of uncertainty in whether to attribute one's behavior to one's own will or to other causes is supported by the many phenomena suggesting an "illusion of conscious will" (Wegner, 2002; Wegner & Wheatley, 1999). For instance, Langer and Roth's (1975) finding that people felt they could control a chance event after correctly predicting a series of like events suggests that knowing whether one is the agent of an event involves an inference, a conclusion that Spanos (1982) also comes to in an effort to explain hypnosis. Wegner and Wheatley take the extreme position that the experience of conscious will "is an experience fabricated from perceiving a causal link between thought and action" (p. 487). Although we cannot rule out the possibility that conscious will has some direct causal properties, the data compel us to agree that people do not have direct access to all actual causes of their behavior.

This gives actors some control over what they learn about themselves from their own actions. If the representation of the causes of one's own action depends on construal, the door is opened to self-deception by giving actors the opportunity to choose the most self-serving representation. Our hypothesis can be interpreted as a specification of Quattrone and Tversky's (1984) hypothesis that "deceptive diagnosis is more likely to occur for actions believed to be uncontrollable than actions believed to be controllable" (p. 243).

To represent an action as an observation is the familiar enterprise of embedding an action in a field of causes and effects, as we normally do when we see a stranger perform some action or we think about our own past actions. But sometimes we need to think about actions as interventions, as deliberate and willful choices. In particular, we need to think of ourselves as free agents who have the ability to change the state of the world. The capacity to represent actions as resulting from interventions is what allows us to experiment, to think about the future, to think about what might have been, and to think about the consequences of potential actions (Hagmayer & Sloman, 2009; Sloman, 2005). Representing an action as the consequence of deliberate choice requires simplifying our causal model in two ways: First, we need to treat the choice itself as uncaused (what probability theorists call *exogenous*). Life if we moved to Florida would be hard to imagine if we get stuck making backwards (diagnostic) inferences about the determinants of choosing to imagine such a life, what the fact that we are thinking about moving would mean about our values, our preferences, and the external pressures on us. Second, we need to assume that the intervention would be effective, that the choice to act would lead to the imagined action. When we imagine moving to Florida, we can just assume we're in Florida. We need not consider all the ways we might not actually arrive there or what arriving there would mean about the availability of flights to Florida. Representing action as the result of deliberate choice means assuming that the action occurred because of an intervention, not for some other reason. The actor becomes fully responsible.

Consider someone who denies an addiction to a drug but actively seeks out the drug at a time when they clearly should not (e.g., having a drink first thing in the morning before going to work). Most observers would take such an action as evidence of addiction; the addiction made them do it (left side of Fig. 1). But an individual who wants to deny an addiction could say they freely chose to consume the drug; the action was not a result of addiction but a result of intervention (right side of Fig. 1). This is a solid argument. One effect of intervention is to make the intervened-on variable independent of its normal causes and thus not diagnostic of them (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993). Self-deception in this case involves representing an action as the result of an intervention when in fact its cause was not agency but addiction in an enabling world. We call this interventional self-deception. Unfortunately, it is not easily recognized because the environment and the agent's preferences both support the same action so there is no way to know with certainty what to attribute it to.

A different kind of self-deception occurs when people treat what is actually the effect of an intervention as an observation. We call this diagnostic self-deception. This is illustrated by a child abuser who claims that external forces (perhaps the child him or herself) made them hit or exploit them. Unless we deny the abuser has any self-control, abuse is an intervention by an agent. Denying responsibility requires the agent to frame his or her own