Cognition xxx (2010) xxx–xxx

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
Available online xxxx

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.

© 2009 Elsevier B.V. All rights reserved.

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

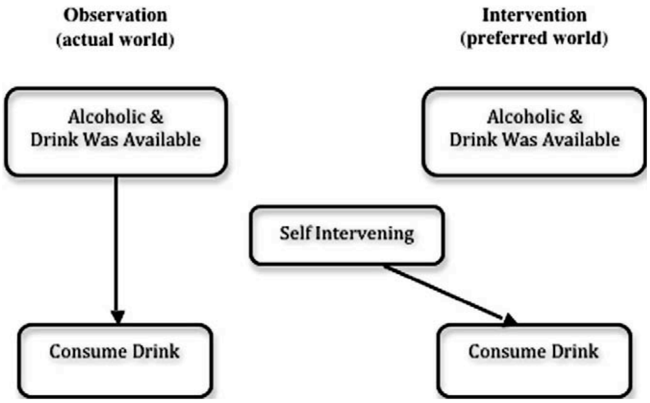

Fig. 1. Self-deception in addiction.

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 behavior as arising from other causes; that they are merely observers of the action.

We are aware of two experimental demonstrations of self-deception and both involve this second kind of self-deception (in both cases, their characterization as "self-deception" is disputed by Mele (1997)). Quattrone and Tversky (1984) asked participants to immerse one hand in very cold water after exercise for as long as they could. Half the participants were told that people who feel a lot of pain from the cold water have a weakness in their cardio-vascular system. This defect leads to early heart attacks and a short life span. The other half was told the opposite, that people who feel little pain have the weakness. The two groups performed differently on the immersion task. The first group held their hands in longer than the second. Presumably this reflected people's desire to have strong cardio-vascular systems. They were treating their pain tolerance and thus the immersion time as evidence about the strength of their hearts. Of course, the fact that a difference obtained between the two groups implies that, overall, they intervened; they manipulated the immersion time. But to maintain the belief that doing so was diagnostic of their hearts, they had to believe they were merely observing their actions (Fig. 2). Participants preferred to believe they were observing themselves when in fact they were intervening. A similar though more complicated analysis applies to an experimental demonstration of self-deception by Gur and Sackeim (1979).

We find this surprising because we have found that people represent the causal structure of their choices with great accuracy (Hagmayer & Sloman, 2009). Specifically, they distinguish intervention and observation and draw corresponding inferences correctly (Sloman & Lagnado, 2005; Waldmann & Hagmayer, 2005). Even rats do (Blaisdell, Sawa, Leising, & Waldmann, 2006). Under what conditions do people represent situations in self-serving ways that defy their actual representational capacities?

We propose that self-deception depends on incomplete knowledge about the environment of choice. Various factors influence choice. People are aware of the influence of some factors, others less so. For instance, people are often unaware that a prime (e.g., Bargh & Chartrand, 1999) or the fluency of a stimulus (Oppenheimer, 2008) affects their choice. Factors that people are aware of and whose influence can be estimated do not lend themselves to self-deception. Once an addict is aware of a craving, it becomes harder for the addict to say to herself that she has no desire for the drug; evidence to the contrary is, by definition, directly available. Self-deception takes advantage of factors whose influence we are not aware of. We can deny that addiction is influencing our behavior only if we are not aware of our cravings. A frequent attribute of factors that influence us without our awareness is that they are not coded explicitly in language or any other representational system. Attributes that are coded vaguely or imprecisely lend themselves to self-deception because they offer alternative interpretations. For example, the line between a craving and a simple preference is not well-specified and this gives an addict the freedom to interpret a craving as a preference thereby preserving the construal of their habit as a choice.

In sum, we propose that diagnostic self-deception entails substituting an observational frame for an interventional one. This substitution imposes two requirements on the self-deceiver: (i) understanding the desirability of manipulating the behavior (this requires causal knowledge) and (ii) the absence of clear feedback that the behavior is manipulated. In other words, one must know at some level the diagnostic value of the variable manipulated (to see the value in manipulating it) while remaining unaware that the variable is being manipulated (to maintain an observational frame). One condition that allows the second requirement to be satisfied is the presence of vagueness in the feedback received about the behavior. When asked to evaluate personal traits, people are more self-serving when those traits refer to ambiguous properties like leadership than unambiguous properties like height (Dunning, Meyerowitz, & Holzberg, 1989; Felson, 1981). We propose that ambiguity of feedback should influence self-deception.

Quattrone and Tversky's (1984) participants might have been able to self-deceive because their choice task had an element of vagueness. The task required a response when

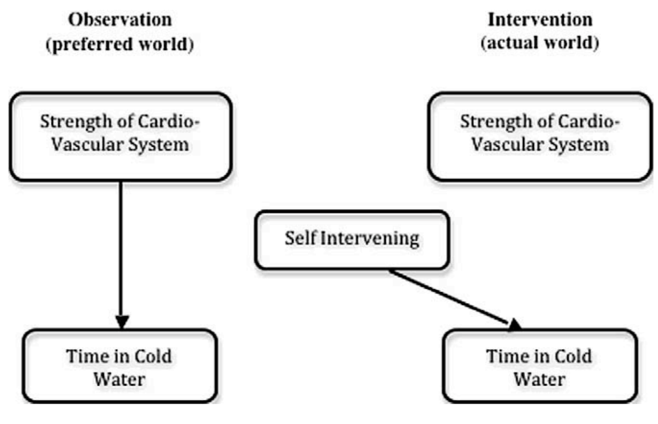

Fig. 2. Self-deception in Quattrone and Tversky (1984). The absence of an arrow from the strength of the cardio-vascular system to time in cold water implies that the intervention was fully determinative. To the extent that the cardio-vascular system were to maintain some influence over and above the intervention, time in cold water should be conceived as jointly determined by both and thus the arrow from the cardio-vascular system would appear on both sides.

pain surpassed some tolerance threshold but the threshold was not (and could not be) articulated precisely. Participants had to react at a point in time of unendurable pain. What is endurable or not is clearly open to interpretation, allowing participants to shift their tolerance while attributing their actions to the pain. If the vagueness had been removed, so that participants had to explicitly take the factor into account when making their choice, we expect that self-deception would not have been observed. But we cannot think of a way to make choice explicit using Quattrone and Tversky's paradigm. Perceptions of pain tolerability are inherently vague.

In our experiments, we used a paradigm with a similar logical structure to Quattrone and Tversky's (1984) study but one that allowed us to change participants' awareness of their influence on their own behavior. We did so by providing participants with feedback. We varied how aware participants were of their intervention on their own performance by varying how precise the feedback was. We predicted that we would only observe self-deception when the feedback was vague enough to allow people to change their behavior in a self-serving direction while attributing the change to some external factor.

2. Experiment 1

The task was a video game in which players started with the cursor on the left side of a computer screen and moved the cursor as fast as possible to a dot that appeared in a random position on the right half of the screen. They were presented with one of two hypotheses about the relation between movement speed and IQ between two phases of dot tracking, an initial and a test phase. In the test phase, they were given vague, qualitative feedback about the speed of their performance.

We hypothesized that difference in performance between the two phases would vary as a function of the hypothesis that was given. Participants would deceive themselves by performing in the test phase in a manner consistent with a self-serving diagnosis of greater intelligence. Participants told that fast performance was linked to high IQ would perform faster than participants told the opposite.

2.1. Methods

2.1.1. Participants and design

Thirty-eight Brown University undergraduates participated for class credit or were paid at a rate of eight dollars per hour. Sessions lasted between 30 and 45 min. Participants were assigned randomly to one of two groups, the fast group or the slow group.

The main independent variable was manipulated between participants. Participants in the fast group were told that fast performance on the tracking task was linked to high general intelligence, while participants in the slow group were told that fast performance was linked to low general intelligence. The only difference between the groups was the wording of three sentences in the background information that participants read about the experiment after completing the tracking task the first time but prior to the test phase.

Two dependent measures were collected, one temporal and one spatial. On each trial the time between the appearance of the dot and the moment at which the participant clicked on the dot was measured. The position of the cursor on the screen was also tracked as each trial unfolded. These spatio-temporal trajectories indicated the position of the cursor on commencement of each trial, and each 100 ms thereafter for 1 s.

2.1.2. Procedure and stimuli

Upon entering the laboratory participants read and signed a consent form and were seated at a computer. They first read on-screen instructions that described the tracking task. The instructions explained that their job was to move the cursor to the dot that appeared on the screen as quickly as they could and that in order to get an accurate reading on each trial they should leave the starting point as soon as the dot appeared, but not earlier. Participants then clicked a button to indicate that they had read and understood the instructions at which point the tracking task screen appeared. After ensuring that the participant understood the task, the experimenter then left the room to allow the participant to complete the tracking task.

The tracking task screen had a button marked with a '+' at the left hand side, equidistant from the top and bottom of the screen. At the right a square border was positioned equidistant from the top and bottom of the screen. There was also a trial counter on the top left of the screen, which indicated the current trial number out of 50.

To initiate each trial, participants clicked the '+' button. One second later a filled red dot appeared at a pre-specified location within the rectangular border. Dot positions for all 50 trials were generated by creating random vectors of *X*- and *Y*-coordinates from a uniform distribution. The same dot positions were used for all participants so that the average distance from the starting point to the dot over all trials was equated. For the test phase, dot positions were generated by randomly permuting the dot position vectors from the first phase. That way the average distance from starting point to dot was also equated across phases of the experiment, but they were presented in different orders. The permutation used in the test phase was the same across all participants. Upon clicking the dot, it disappeared indicating the end of the trial. Participants then returned to the starting point and clicked the '+' button to commence the next trial, continuing until all 50 trials were completed.

After completing the tracking task, participants were given a sheet marked 'Background Information' to read. This sheet described the purpose of the experiment and described the link between performance on the tracking task and general intelligence. Participants in the fast group read:

> This experiment is aimed at looking at different types of computational speed and their relationship to each other. Computational speed is the ability to process information rapidly. The tracking task that you just performed is used to measure a special kind of computa-

> tional speed, called spatio-motor speed. Spatio-motor speed is the ability to rapidly locate and access spatial locations. *Many recent experiments have shown that spatio-motor speed is linked to higher general intelligence. People who perform well on tracking tasks like the one you just completed tend to have higher than average general IQ scores. It is believed that the spatio-motor system and higher cognitive capacities share underlying computational processes which cause people with high spatio-motor speed to also have high general intelligence* (italics added).
>
> A second kind of computational speed associated with higher cognitive function is called processing speed. Processing speed is the ability to rapidly perform mental computations and access information. In this experiment we are assessing whether there is a link between spatio-motor speed and processing speed by looking at how performing mental computations and information processing tasks prior to the tracking task affects performance.

In the slow group, the italicized part of the first paragraph read:

> Many recent experiments have shown that spatio-motor speed is linked to lower general intelligence. People who perform well on tracking tasks like the one you just completed tend to have lower than average general IQ scores. It is believed that the spatio-motor system competes with higher cognitive capacities for computational resources and decreases overall performance on tests measuring general IQ.

After reading this information, participants answered two questions; how likely they thought it was that they had above-average spatio-motor speed, and how happy they would be if they found out that they had above-average spatio-motor speed. Responses to both were given on a one-to-ten scale.

Next they read about the 'plan for the experiment.' They were told that they would first 'perform a control task to return your computational system to baseline levels,' then complete a battery of processing speed tasks, and finally complete the tracking task again. The purpose of the control and processing speed tasks was to increase participants credulity that we were actually testing what we claimed to be testing and to increase the plausibility of the link between the tracking task and general intelligence.

After reading the background information, participants returned to the computer to complete the control task of 40 simple estimation questions. After the baseline task the participants were seated across from the experimenter to complete the processing speed tasks. There were five of these, each lasting 1 min. Participants were instructed to try to give as many answers as possible in the allotted time, while trying to be as accurate as possible. The experimenter wrote down the answers as they were given on a sheet of paper that was not visible to the participant. The five tasks in order of completion were to (a) generate the Fibonacci sequence starting at eight, (b) generate the sequence X^2 starting at $X = 4$, (c) generate the sequence 2^X starting at $X = 4$, (d) name as many capital cities of countries as possible, and (e) list as many words as possible of three letters or more that can be spelled by using letters from the word 'environment.'

After completing the processing speed tasks participants returned to the computer to complete the test phase of the tracking task. Participants were told that the only difference from the first phase of the tracking task was that they would now receive feedback on their performance. They read the tracking task instructions again, which were identical to the first phase except they reiterated that participants would receive feedback. The procedure for the trials was the same as the first phase. The screen differed only in that a horizontal red bar above the square border indicated the participant's speed on the last trial. After clicking on the dot, a vertical yellow line appeared within the red bar and the word 'fast', 'slow' or 'average' was displayed above the bar. An example screenshot of the feedback is shown in Fig. 3. Performance on the trial was determined by comparing the time for the trial to the times from the

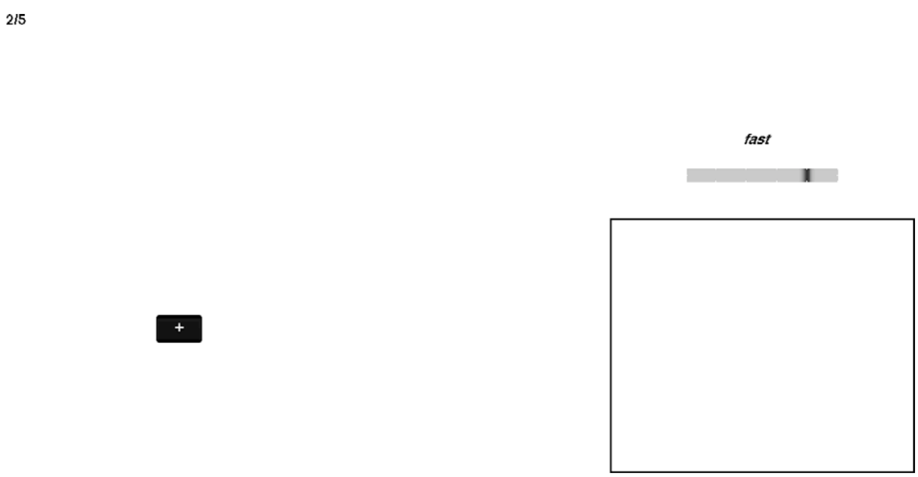

Fig. 3. Screen shot from test phase of Experiment 1 rendered in black and white. Actual experiment was in color.

Please cite this article in press as: Sloman, S. A., et al. Self-deception requires vagueness. *Cognition* (2010), doi:10.1016/j.cognition.2009.12.017

first phase. The position of the yellow line within the bar represented the percentile of the current time, with percentile decreasing from left to right. 'Fast' was displayed if the time was in the lower 30% of times from the first phase, 'slow' if the time was in the upper 30%, and 'average' if it was in the middle 40%. Participants were not told what the feedback meant. They simply saw the position of the yellow line in the red bar, and whether the trial was 'fast', 'slow' or 'average.'

After completing the 50 trials in the test phase participants answered three more questions. First, they were asked again on a one-to-ten scale how likely it was that they had above-average spatio-motor speed. Next they were asked if the information they read about computational speed and IQ affected how they performed the tracking task, and if so, how. Lastly they were asked if they 'cheated' by purposely moving the cursor before or after the dot appeared.

Finally, they were told that the link between IQ and computational speed was fabricated for the experiment. They were then asked if they had suspected the links were fabricated and if so how confident they were on a one-to-ten scale. After this they completed the experimental session by reading a paragraph explaining the true purpose of the experiment.

2.2. Results

Four participants indicated that they changed their performance in the test phase in response to the cover story. All four came from the fast group and claimed to have intentionally speeded up in response to learning that fast performance was linked to high general intelligence. These participants are not included in the subsequent analyses. No participants admitted to cheating by purposely moving the cursor significantly before or after the dot appeared.

2.2.1. Analysis of speed

Outliers were removed from the dataset by eliminating trials whose time was more than three standard deviations from the mean of that participant's performance in that phase. In total, 45 data points were removed, amounting to 1.2% of the dataset. The maximum number of outliers for any subject was 3.

Mean trial times and standard errors are shown in Fig. 4. For the first phase, participants in both fast and slow groups required 1.11 s; they did not differ, $t(32) < 1$ (all statistical tests on reaction times in this paper used a log transformation to normalize the data). Some learning was observed as speeds were faster in the second than the first phase. Means for the test phase were 0.99 s and 1.05 s for the fast and slow conditions, and these were significantly different, $t(32) = 2.06$, $p < .05$. Most relevant to our hypothesis, speedup from initial to test phase was greater in the fast condition (difference of .12 s) than the slow condition (.07 s), $t(32) = 4.47$; $p < .0001$. We refer to this differential speedup as self-deception because it indicates that participants manipulated their behavior even though such manipulation was self-defeating. It prevented them from learning about the property that motivated the behavior in the first place, their computational speed.

2.2.2. Trajectories

We took measurements of cursor position at the commencement of each trial and 100 ms thereafter nine more times, for a total of 10 data points for each trial. For each participant we then calculated a mean position for each time point for the 50 trials in the first phase and the 50 trials in the test phase.

If participants 'cheated' to improve or diminish performance we would have expected different trajectories between phases and between groups. If participants in the fast group left the dot early in the test phase to go faster, we would expect to see the position of the cursor further to the right of the screen on commencement of the trial relative to the first phase and relative to the slow group. Analogously, if participants in the slow group hesitated in order to diminish performance, their cursor positions would have remained leftward for longer relative to the first phase and the fast group.

Average trajectories for the first phase and the test phase for each group are shown in Fig. 5. For simplicity only progress in the x-direction (i.e. left to right across the screen) is depicted. The x-axis represents time as it unfolds over the course of the trial. The y-axis represents cursor position as a percentage of total progress toward the target. Thus, 0% on the y-axis means that the cursor was positioned at the starting point, and 100% means that the cursor reached the target. The average position late in the trial is slightly above 100% because participants sometimes overshot the target.

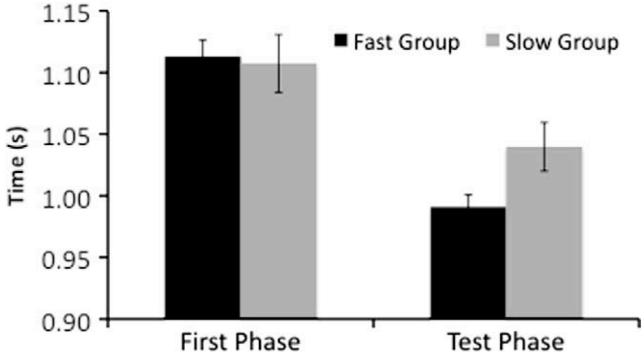

Fig. 4. Mean reaction times with standard error bars for Fast and Slow Hypothesis groups in both Initial and Test phases of Experiment 1.

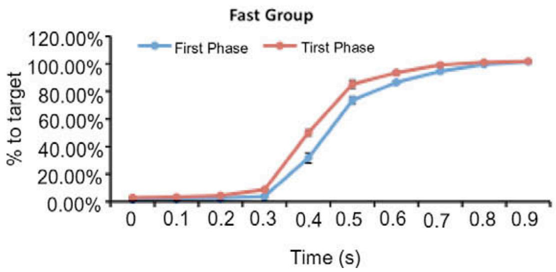

Fig. 5. Mean trajectories for Fast and Slow Hypothesis groups in both Initial and Test phases of Experiment 1.

Visual inspection of the trajectories shows no evidence of cheating; participants left the starting point at the same point in the trial in the first phase and the test phase. The trajectories do look different across phases because speed increased from the first phase to the test phase, and the difference is more pronounced for the fast group than the slow group. However, the difference is not due to participants leaving the starting point early or late depending on condition. The difference is due to greater acceleration in the fast group after leaving the starting point, beginning between 300 and 400 ms after the appearance of the dot.

We checked those trials on which participants left the starting point early to see if they were aware of having done so and corrected by moving the cursor back toward the starting point. Participants left the starting point early on 80 of the 3800 trials in the experiment (2.1%). On those 80 trials participants moved the cursor back toward the left later in the trial ('backtracked') 41 times (51.3%). For comparison, on trials in which they did not leave the starting point early, participants backtracked on just 1.8% of trials, a significantly lower proportion ($Z = 24.3$, $p < .0001$). When participants did leave the starting point early, they seem to have been aware of doing so.

2.2.3. Inference

Participants were asked how happy they would be if they found out they had above-average spatio-motor speed after reading about the link between speed and general intelligence. Mean responses for the fast group and the slow group were 6.59 (SE = .29) and 6.00 (SE = .54) respectively, not significantly different, $t(32) < 1$. A difference may not have arisen here because a large proportion of participants in both groups gave responses of five or six, the midpoints of the scale. Also, a small number of participants in the slow group unexpectedly gave very high responses. A typical justification was "I know I'm smart, and I would be very happy to find out that I'm also fast." Responses to this question were not significantly correlated with the magnitude of their speed up in the test phase of the tracking task ($r = 0.26$, *n.s.*).

Participants were also asked the likelihood that they had above-average spatio-motor speed both after reading about the link between spatio-motor speed and IQ but before completing the test phase, and after completing the test phase. Before the test phase, the mean across participants in the fast group was 6.35 (SE = .32), marginally higher than the slow group's 5.29 (SE = .44), $t(32) = 1.94$, $p = .06$. After the test phase, the fast group's mean was identical (6.35; SE = .32), even closer to the slow group (5.88; SE = .43). The mean difference in response after completing the test phase was 0 (SE = .21) for the fast group and .59 (SE = .26) for the slow group, $t(32) = 1.77$, $p = .09$.

The lack of difference between the two questions could reflect a desire to be consistent (responses to the two questions were highly correlated, .81). This would have prevented a difference from arising between groups. It would also explain the lack of correlation between their answers to the second question and their speedup in reaction times (.14; *n.s.*).

2.2.4. Plausibility of manipulation

Eight of the 17 participants in the fast group and 12 of 17 in the slow group said that they had suspected that the link between speed on the tracking task and general intelligence was fabricated. The fact that so many were skeptical about the hypotheses makes the finding of self-deception even more surprising.

To determine how beliefs about the credibility of the hypotheses influenced self-deception, we compared the amount of speedup on the tracking task for the 15 participants who believed that the hypotheses were fabricated and said so with confidence (their confidence rating was above the midpoint of the scale) to the amount of speedup exhibited by the remaining 19 participants. Only the second group showed self-deception (Fig. 6). Speedup between the fast and slow groups did not differ for those participants who were confident of fabrication, $t(13) < 1$. It did differ for the remaining participants, $t(17) = 4.31$; $p < .001$. As would be expected, self-deception only occurred for those participants who considered our hypotheses credible.

2.2.5. Processing speed

The two groups did not differ on any of the measures of processing speed administered between the two phases. Some of the measures correlated with each other but none showed significant correlations with our measure of self-deception, the amount of speedup between the two phases.

2.3. Discussion

Participants who were told that fast performance on the tracking task was linked to high general intelligence in-

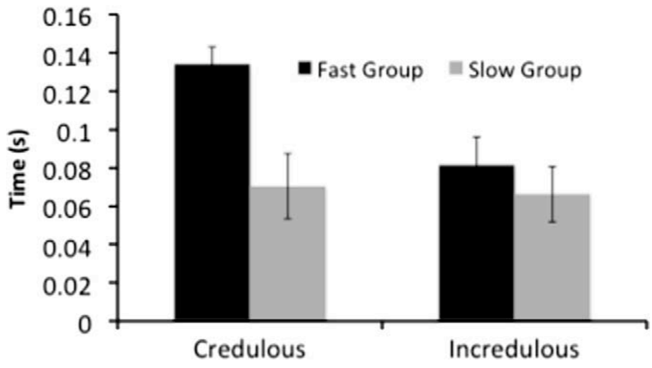

Fig. 6. Mean difference in reaction times between the first phase and the test phase with standard error bars for Fast and Slow Hypothesis groups as a function of belief that the data were fabricated Experiment 1.

creased their speed in the test phase of the task more than participants who were told the opposite despite claiming that they did not alter their performance. Participants deceived themselves, generating responses correlated with a desired attribute, even though the decision to generate those data should have rendered them non-diagnostic of intelligence.

This experiment merely replicates Quattrone and Tversky (1984) and Gur and Sackeim (1979) in demonstrating self-deception. However, it also provides a hint in support of our main hypothesis that self-deception will only occur in the face of vagueness. Participants did not 'cheat' by leaving the starting point early or late to augment or diminish performance. This would have been too obvious, and in fact, participants did tend to notice when they left early and corrected by backtracking. They deceived themselves by speeding up, an act, like holding one's hand in cold water, which does not leave an explicit trace available to conscious awareness. One can speed up and remain unaware of doing so (as we all do frequently when we match the gait of someone we're walking with or even talking to on a cell phone, Murray-Smith, Ramsay, Garrod, Jackson, & Musizza, 2007).

3. Experiment 2

Experiment 2 was identical to Experiment 1 except that participants were provided with more precise feedback in the test phase. We predicted that greater precision in the information participants had about their behavior would reduce the incidence of self-deception.

3.1. Methods

Methods were identical to Experiment 1 in all respects except the following: 36 participants were tested and the feedback in the test phase was more explicit. In Experiment 1 participants saw a slider moving up along a single dimension along with the word 'fast, 'slow' or 'average' but no context to interpret the feedback. In the Experiment 2 test phase, participants were shown a scatter plot above the square border, which had their reaction time on each trial plotted as a small red dot. After each trial, the new data point appeared on the chart and flashed three times to draw attention to it. All of the times from the first phase of the experiment were also on the chart on a line below

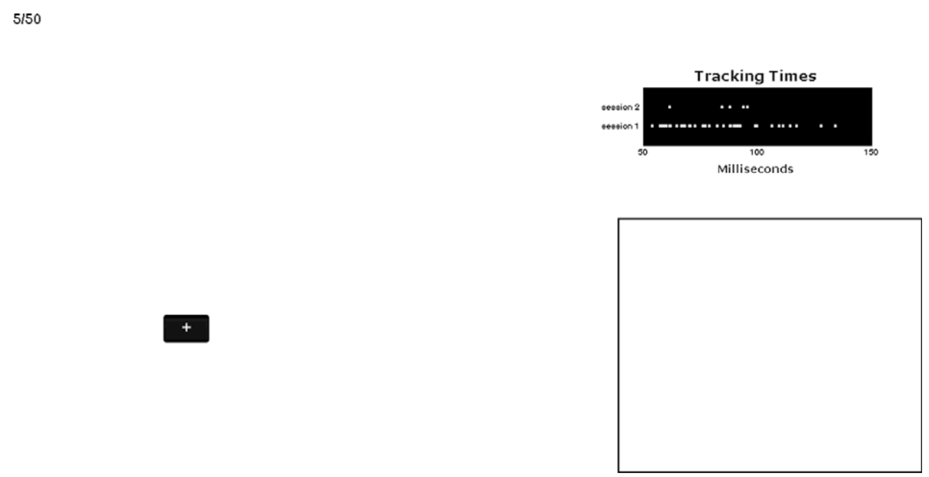

Fig. 7. Screen shot from test phase of Experiment 2 rendered in black and white. Actual experiment was in color.

Please cite this article in press as: Sloman, S. A., et al. Self-deception requires vagueness. *Cognition* (2010), doi:10.1016/j.cognition.2009.12.017

the times from the test phase. A screenshot of the display as it appeared during the test phase is shown in Fig. 7. Prior to beginning the test phase, the experimenter described how to interpret the chart and clearly stated what the dots on the chart represented.

3.2. Results

Six participants (all from the fast group) were excluded from the subsequent analyses because they indicated that they changed their performance in the test phase in response to the cover story. Again, no one admitted to cheating by moving the cursor early or late.

3.2.1. Analysis of speed

Outliers were again removed from the dataset. This time, 45 data points were removed (1.2%). The maximum number of outliers for any participant was 3.

Mean trial times and standard errors are shown in Fig. 8. For the initial phase, participants in fast and slow groups required 1.10 and 1.13 s, respectively (SEs were .014 and .025), $t(28) < 1$ comparing groups. Again there was learning: Means for the test phase were 1.03 s and 1.065 s for the fast and slow conditions, respectively. This time, the groups did not differ, $t(28) = 1.17$, *n.s.* We did not observe self-deception: Speedup from initial to test phase did not differ (.07 s in the fast condition and .06 in the slow), $t(28) < 1$.

The data are consistent with our prediction of a greater speedup in the fast than slow conditions of Experiment 1 but not Experiment 2. To test this prediction directly, we ran a 2 × 2 analysis of variance that included Experiments 1 versus 2 as an independent variable along with Hypothesis (Fast versus Slow). The interaction was significant ($F(1,58) = 4.21$, $p < .05$) as were the main effects of Experiment, $F(1,56) = 5.86$; $p < .05$, and Hypothesis, $F(1,56) = 8.48$, $p < .01$. Means and SEs are shown in Fig. 9.

3.2.2. Trajectories

As expected, the trajectories for the fast and slow groups did not differ.

3.2.3. Inference

Unlike Experiment 1, the two groups provided different ratings of how happy they would be if they found out they had above-average spatio-motor speed. Mean responses for the fast group and the slow group were 6.93 (SE = .30) and 5.33 (SE = .42) respectively, $t(28) = 3.09$, $p < .01$. However, these responses were again uncorrelated with the magnitude of their speedup in the test phase of the tracking task ($r = -.19$, *n.s.*).

Inferences about spatio-motor speed did not differ. Before the test phase, the mean in the fast group was 6.00 (SE = .37) and in the slow group's 5.6 (SE = .51). After the test phase, both groups' means actually went down a little to 5.13 (SEs = .36 and .46 in the two groups, respectively).

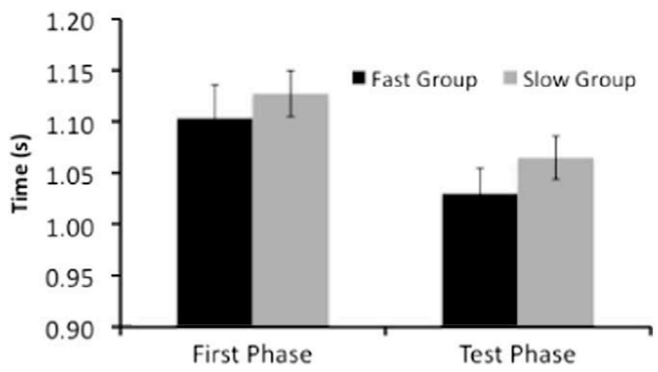

Fig. 8. Mean reaction times with standard error bars for Fast and Slow Hypothesis groups in both Initial and Test phases of Experiment 2.

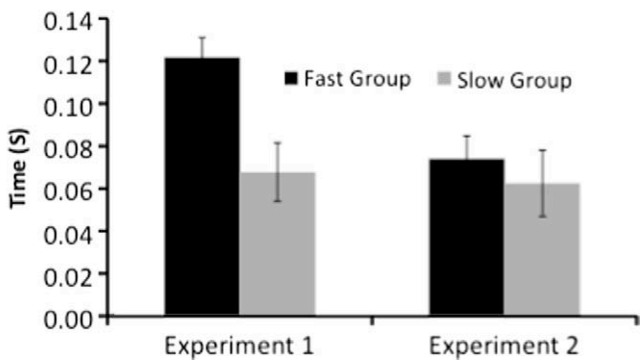

Fig. 9. Mean reaction time differences between Initial and Test phases with standard error bars for Fast and Slow Hypothesis groups in Experiments 1 and 2.

The amount of change did not differ (.87 and .47, respectively), $t(28) = 1.11$, *n.s.* Again, there was no correlation between their answers to the second question and their speedup in reaction times (.08; *n.s.*).

3.2.4. Plausibility of manipulation

Twelve of 21 participants in the fast group and 11 of 15 in the slow group said that the hypotheses might be fabricated. In this experiment, no subgroup evidenced self-deception regardless of their beliefs about the evidence.

3.2.5. Processing speed

Again the groups did not differ on any of the measures of processing speed. Some of the measures correlated with each other but not with the amount of speedup between the two phases.

3.3. Discussion

Experiment 2 differed from Experiment 1 in only one respect, the nature of the feedback given during the tracking task. Unlike the vague feedback of Experiment 1, Experiment 2 provided explicit and quite precise feedback on each trial that revealed how performance compared to all other trials in the experiment. This minimal change was enough to eliminate the difference between participants who had been given a hypothesis suggesting going fast was desirable versus those given one that favored going slowly.

One way to describe what happened in Experiment 2 is that the fast group converged to the slow group. That is, performance in Experiment 1's slow group was almost identical to performance in both groups of Experiment 2. This suggests that self-deception only occurred in the fast group. It may be that people can deceive themselves about trying to go faster than they otherwise would, but not about trying to go slower. Alternatively, judgments about whether the hypotheses were fabricated suggest that the slow hypothesis was less plausible than the fast one. People may not deceive themselves in favor of a hypothesis that they do not find credible.

4. Experiment 3

The interaction between the vagueness of feedback and the fast/slow instructions we observed comparing Experiments 1 and 2 supported the hypothesis that self-deception is attenuated in the presence of explicit feedback. In this experiment, we tested the influence of the form of feedback within a single experiment. Because fewer participants believed our slow hypothesis, and for reasons of economy, we only used Fast instructions. Comparison of Experiments 1 and 2 suggested that self-deception only occurred with Fast instructions so any self-deception should show up in that condition. Feedback was manipulated between participants. The Implicit group received feedback as in Experiment 1, while the Explicit group received feedback as in Experiment 2.

The other goal of Experiment 3 was to further explore the extent to which people are aware of their own influence on their behavior. Quattrone and Tversky (1984) and Experiments 1 and 2 used a dichotomous yes/no response to determine awareness of influence. This leaves open the possibility that if people had intermediate options they might have expressed some level of awareness. Gur and Sackeim (1979) used a galvanic skin response measure as a non-verbal measure but subsequent research contested the validity of this measure (Douglas & Gibbins, 1983). In Experiment 3 we gave participants the opportunity to express degrees of uncertainty about whether they changed their performance by using a graded scale. If participants still indicate no influence, then this provides stronger evidence for lack of awareness.

4.1. Methods

Forty introductory psychology students participated for course credit and were assigned randomly to the Explicit and Implicit conditions. Methods were identical to the Fast conditions from Experiments 1 and 2, respectively, except that a 7-point scale was used to answer the question about whether performance had been altered in response to the cover story. The response options from left to right were, "I Definitely Did Not", "No, I did not", "Probably Not", "Maybe", "Probably", "Yes I Did," and "I Definitely Did". Also, the question about spatio-motor speed was only asked after the test phase.

4.2. Results

4.2.1. Analysis of speed

Analyses include only participants who said they "definitely did not" or "did not" change their performance. In the Explicit group, five answered 'definitely did not' and five answered 'did not'. In the Implicit group, one answered 'definitely did not' and nine answered 'did not'. The tendency to avoid an extreme response may reflect that participants were psychology students and most were aware of motivational effects in cognition. The remaining participants (eight from the Explicit group and 12 from the Implicit group) gave a different response and were eliminated from this analysis. The proportions of participants excluded did not differ across the conditions ($Z < 1$, *n.s.*).

Fifty-six outliers were removed (1.4%) from the data set using the same criteria as previous experiments. The maximum number for any individual was 3. Mean trial times and standard errors are shown in Fig. 10. Corroborating the interaction between Experiments 1 and 2, speedup from the initial to the test phase was significantly greater for the Implicit group (.12 s) than the Explicit group (.06 s), $t(18) = 3.9$, $p < .01$. For the initial phase, participants in the Implicit and Explicit groups required 1.07 and 1.14 s, respectively (SEs were .048 and .031), $t(18) = 1.4$, $p = .19$. For the test phase, they required 0.95 and 1.08 s, respectively (SEs were .039 and .033), $t(18) = 2.7$, $p < .05$.

Four participants answered they "probably did not" change their performance. Post-experiment interviews indicated that they had no conscious awareness of changing performance but were willing to concede that it was possible that they could have been influenced subcon-

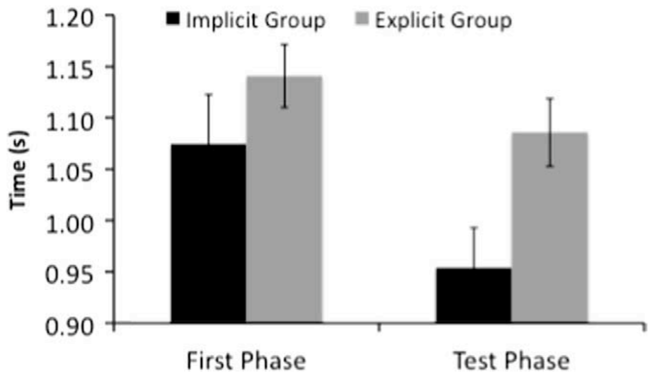

Fig. 10. Mean reaction times with standard error bars for Implicit and Explicit groups in both First and Test phases of Experiment 4.

sciously. Including the data from this group does not affect the results: speedup remains significantly greater for the Implicit group (.12 s) than the Explicit group (.05 s), $t(22) = 3.9$, $p < .001$.

4.2.2. Trajectories

Again, there was no evidence of cheating by leaving the starting point early.

4.2.3. Inference

Both groups gave ratings above the scale midpoint of their happiness if they were to find out they had above-average spatio-motor speed. Means were 6.6 (SE = .39) and 6.8 (SE = .37) for the Implicit and Explicit groups, respectively. The correlation between happiness judgments and the magnitude of speedup was positive but not significant, $r = .24$, $p = .32$.

Inferences about spatio-motor speed were in the same range as the previous experiments. The mean for the Implicit group (6.1, SE = .56) was higher than the Explicit group (5.5, SE = .50) but not significantly so, $t(18) = .7$, $p = .46$. Again, the correlation between inferences about spatio-motor speed and speedup in reaction times was not significant, $r = -.17$, $p = .46$.

4.2.4. Plausibility of manipulation

Nine of 18 participants in the Implicit group and 10 of 22 in the Explicit group said that they thought the hypotheses might be fabricated with confidence above the scale midpoint. To increase statistical power comparing credulous to incredulous participants, we included the participants who said they 'probably did not' change their performance. Greater speedup was observed for the Implicit group among both credulous, $t(8) = 3.8$, $p < .01$, and incredulous, $t(12) = 2.2$, $p < .05$, participants though the effect was larger for the credulous ones. Among participants who said 'maybe', 'probably I did' or 'I definitely did', 5 of 16 were confident of fabrication.

4.2.5. Processing speed

Performance on the processing speed tasks was similar to the other experiments. Again, none of the measures correlated with the amount of self-deception.

4.3. Discussion

Experiment 3 replicates the difference found comparing Experiments 1 and 2 providing further evidence that self-deception depends on vagueness in the information available about performance. Vagueness in feedback gives people leeway to attribute their performance to their own intervention or to spatio-motor speed. However, the feedback differed not only in its vagueness, but in another way that could potentially offer an alternative account of the results. Explicit feedback was not only more precise, reaction times specified to the millisecond, it also explicitly concerned each participant's own performance and thus gave no information about how fast participants were relative to others and thus no information relevant to the hypothesis about intelligence that they were presented with. In contrast, Implicit feedback was unclear in both the precise meaning of the terms used ('fast', 'slow,' or 'average') and the reference class. Were times fast relative to participants' own performance, some group's performance, or to some other standard?

5. Experiment 4

Experiment 4 was designed to determine whether the critical difference in Implicit versus Explicit feedback concerned the precision of the terms or the clarity of the reference class. The methods were again identical to Experiment 1 except that the reference group used to determine the feedback was specified. In one condition, participants were told that the feedback was relative to their own performance (the Self group). In a second condition, they were told that it was relative to a group of other Ivy League students (the Population group). If self-deception depends only on the presence of vagueness in feedback, then self-deception should be observed in both conditions. If it depends on vagueness in the reference class of the evaluation, then it should not be observed in either.

The experiment also allows us to examine whether self-deception requires that data be diagnostic. The hypotheses that participants were given implied that the speed of their performance was related to how fast they went relative to other people. Therefore, only feedback that described their performance relative to other people was diagnostic of their intelligence, not feedback that described their performance relative to themselves. If self-deception is sensitive to diagnosticity, we should observe self-deception in the Population but not the Self condition.

Like Experiment 3, we only used the fast instructions. Conditions that did not show self-deception can serve as a baseline. We chose to use Experiment 1's Slow condition as a baseline because it is the most directly comparable procedurally.

5.1. Methods

Methods were identical to Experiment 1 in all respects (including the use of vague speed adjectives) except the following: 56 participants were tested and the reference class used to determine feedback in the test phase was made explicit. Prior to beginning the test phase the experimenter explained to participants that they would be receiving feedback. In the Self group they were told that the feedback was comparing trial times to their performance in the first phase while the Population group was told that it was comparing their performance to a group of other Ivy League students who had previously performed the task. Also, the question about computational speed was only asked after the test phase to test whether the lack of change in people's beliefs about their speed was driven by a desire to give consistent answers to the question.

5.2. Results

Thirteen participants (3 from the Self and 10 from the Population group) were eliminated because they indicated that they changed their performance in the test phase in response to the cover story. Again, no one admitted to cheating by moving the cursor too early or too late.

5.2.1. Analysis of speed

Outliers were again removed from the dataset. Sixty-one data points were removed (1.1%). The maximum number for any subject was 4.

Mean trial times and standard errors are shown in Fig. 11. For the initial phase, participants in the Self and Population groups required 1.12 and 1.13 s, respectively (SEs were .016 and .036), $t(28) < 1$. As before, means for the test phase were faster, 1.03 s and 1.02 s, respectively (SEs were .017 and .030). The groups did not differ, $t(28) < 1$, neither did their degree of speedup from initial to test phase (.10 s in the Self condition and .11 in the Population), $t(28) < 1$.

To measure self-deception, we compared speedup (the difference between the Initial and Test phases) in each condition to speedup in Experiment 1's Slow condition. Self-deception was observed in both conditions (see Figs. 11 and 12). Speedup in the Self condition (.10 s, SE = .012) was greater than in the Slow condition (.07, SE = .011), $t(37) = 2.05$; $p < .05$ as was speedup in the Population condition (.11 s, SE = .012), $t(34) = 2.20$; $p < .05$.

5.2.2. Trajectories

The trajectories for both groups evidenced the same pattern as the fast group of Experiment 1 with greater acceleration in the test phase than the first phase beginning 300–400 ms into the trial. Again, there was no evidence of cheating by leaving the starting point early.

5.2.3. Inference

Both groups gave high ratings of their happiness if they were to find out they had above-average spatio-motor speed. Means were 7.22 (SE = .17) and 7.26 (SE = .34) for the Self and Population groups respectively. This was higher than the ratings in the Slow condition of Experiment 1 for the Self group, $t(37) = 2.4$; $p < .05$, and for the Population group, $t(34) = 2.03$; $p < .05$. However, happiness judgments were again uncorrelated with the magnitude of their speedup in the test phase, $r = -.10$, *n.s.*

Inferences about their spatio-motor speed were in the same range as in Experiments 1–3. The mean was 5.68 (SE = .26) in the Self condition and 6.32 (SE = .43) in the Population condition. Again, there was no correlation between their answers to the second question and their speedup in reaction times (−.01; *n.s.*).

5.2.4. Plausibility of manipulation

Fourteen of 25 participants in the Self group and 16 of 29 in the Population group said that they thought the hypotheses might be fabricated. In the Self condition, that subgroup that was confident that the data were fabricated

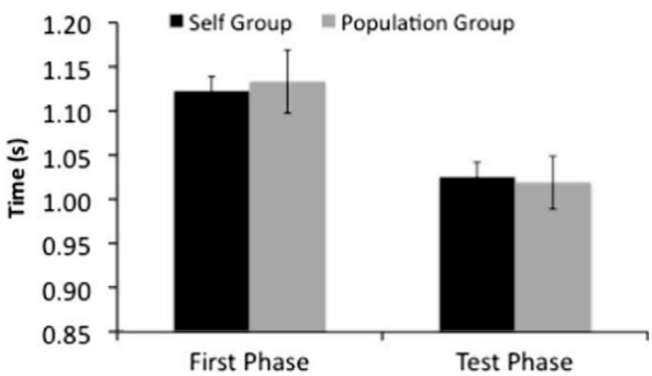

Fig. 11. Mean reaction times with standard error bars for Self and Population Feedback groups in both Initial and Test phases of Experiment 3.

Please cite this article in press as: Sloman, S. A., et al. Self-deception requires vagueness. *Cognition* (2010), doi:10.1016/j.cognition.2009.12.017

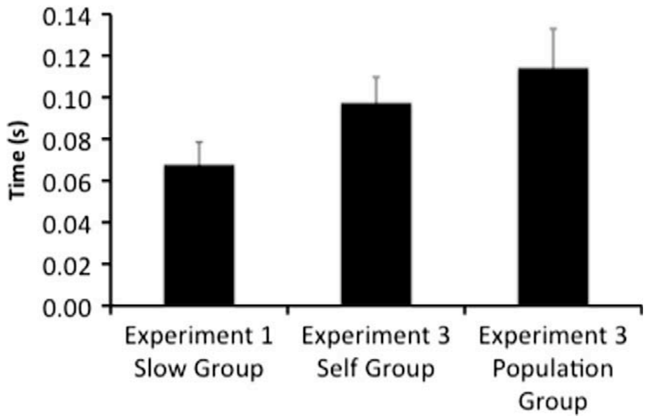

Fig. 12. Mean reaction time differences between Initial and Test phases with standard error bars for Slow Hypothesis group in Experiment 1 and Self and Population Feedback groups in Experiment 4.

had a mean reaction time difference of .11, greater than the corresponding group in Experiment 1's Slow condition (.068), $t(15) = 2.3$; $p < .05$. In the Population condition, that subgroup had a mean of .10, also greater than the same comparison group but not significantly, $t(18) < 1$. In sum, the means for both groups were in a direction suggesting self-deception but only the Self group comparison was significant. The limited sample sizes for these comparisons prevent any strong conclusion.

5.2.5. Processing speed

Performance on the processing speed tasks was similar to the other experiments. Again, none of the measures correlated with the amount of self-deception.

5.3. Discussion

In Experiment 4, vague feedback was used. Self-deception was observed regardless of the reference class of the feedback. All participants were told that going fast was related to intelligence. Speedup in the test phase occurred relative to the slow condition of Experiment 1 both when feedback was relative to their own performance and thus not diagnostic, or when it was diagnostic, relative to a group of other Ivy League students.

6. General discussion

Self-deception occurred in our dot-tracking task in the form of speeding up by people who were told that those who go faster are more intelligent (Experiments 1, 3, and 4). Self-deception only occurred however when feedback on each trial was vague, when it was presented in qualitative terms that afforded a self-serving construal (implied by Experiments 2 and 3).

Self-deception apparently has two requirements. First, the task must involve an imprecise response that gives participants the opportunity to put out an extra effort. This would explain why speedup in Experiment 2, in the absence of self-deception, was equal to speedup in the slow condition of Experiment 1. We interpret this as evidence that self-deception was only observed when participants were told that going faster was correlated with intelligence (fast condition of Experiment 1, Implicit condition of Experiment 3, and both conditions of Experiment 4). One can go faster by putting out an extra effort. Because participants were supposed to be trying hard anyway, they could explain away the extra effort as simply performing the task. But going more slowly than one is able requires doing something opposed to task requirements and is not as easy to hide from oneself.

Second, self-deception requires imprecise feedback on performance. Simply having vague terms ('fast', 'slow,' or 'average') is sufficient to satisfy this requirement. Precise feedback makes it too obvious that the agent is intervening. The specific reference class used to determine the classification does not matter (Experiment 4). It can be diagnostic by referring to a broader population or it can be non-diagnostic by referring to speed relative to the individual's own earlier performance, a form of feedback irrelevant to the hypotheses about intelligence.

Experiment 1 offered a preliminary suggestion that self-deception does not occur when participants do not believe the target hypothesis. But this result was not replicated in either Experiments 3 or 4; Implicit feedback always led to self-deception.

We saw little evidence that people actually learned about themselves through self-deception. In conditions exhibiting self-deception they did not assert a significantly higher likelihood of having above-average speed than in conditions not exhibiting self-deception. Quattrone and Tversky (1984) found that participants who self-deceived in that they denied adjusting their performance were more likely to draw inferences about their hearts after the cold-water task than participants who did not deny adjusting their performance. No strong conclusion about willingness to make explicit inferences about oneself can be drawn from the Quattrone and Tversky analysis because it suffers from a possible participant-selection bias: Deniers might be more willing to draw inferences from behavior regardless of performance.

The reason for the absence of learning in our Experiment 1 could be that participants were asked about their speed twice, the first time before deceiving themselves, and they may have desired to give a consistent response the second time. It is not clear to us why a more positive inference was not drawn in Experiments 3 or 4. Perhaps it was too obvious and thus would have revealed the deception. That is, the failure to draw the inference explicitly may itself have been a form of self-deception. Another possibility concerns the nature of the attribute people were deceiving themselves about. Unlike heart type, participants already know something about their intelligence. It may be that self-deception in our experiments merely served to confirm what they already believed.

Like previous demonstrations (Gur & Sackeim, 1979; Quattrone & Tversky, 1984), self-deception in our experiments involved participants convincing themselves they were merely observing when in fact they were intervening. We have concluded that self-deception in such cases requires a task demanding an imprecise response whose performance can be improved though effort and that feedback on performance be imprecise. These conditions permit people the greatest leverage in convincing themselves that their behavior is a result of their abilities and not their free will. Other kinds of behaviors, like addiction, entail a different form of self-deception, convincing oneself that action is a function of free will and not physical, psychological, and emotional need. Such self-deception is likely to impose different requirements.

A further requirement that applies to both cases is an accurate model of the causal forces impinging on the agent. In order to deceive oneself about the causes of action, one must understand at some level what the possible causes are. People have sophisticated representations and inference mechanisms for causal reasoning, especially for reasoning about behavior. Diagnostic self-deception does not subvert these mechanisms but rather takes advantage of them.

Acknowledgments

This work was funded by NSF Award 0518147. We thank Jonathan Bogard for collecting data and Jonathan, Marc Buehner, Adam Darlow, Ju-Hwa Park, Michael Ranney, and John Santini for their comments and ideas during the course of this research.

References

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462–479.

Baumeister, R. F. (1993). Lying to yourself: The enigma of self-deception. In M. Lewis & C. Saarni (Eds.), *Lying and deception in everyday life* (pp. 166–183). Guilford Press.

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183–200.

Blaisdell, A. P., Sawa, K., Leising, K. J., & Waldmann, M. R. (2006). Causal reasoning in rats. *Science, 311*, 1020–1022.

Dawson, E., Gilovich, T., & Regan, D. T. (2002). Motivated reasoning and performance on the Wason selection task. *Personality and Social Psychology Bulletin, 28*, 1379–1387.

Douglas, W., & Gibbins, K. (1983). Inadequacy of voice recognition as a demonstration of self-deception. *Journal of Personality and Social Psychology, 44*(3), 589–592.

Dunning, D. (1999). A newer look: Motivated social cognition and the schematic representation of social concepts. *Psychological Inquiry, 10*, 1–11.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*, 1082–1090.

Felson, R. B. (1981). Ambiguity and bias in the self-concept. *Social Psychology Quarterly, 44*, 64–69.

Goleman, D. (1985). *Vital lies, simple truths: The psychology of self-deception*. New York: Simon & Schuster.

Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology, 37*, 147–169.

Hagmayer, Y., & Sloman, S. A. (2009). Decision makers conceive of their choice as intervention. *Journal of Experimental Psychology: General, 138*, 22–38.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*, 636–647.

Langer, E. J., & Roth, J. (1975). Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task. *Journal of Personality and Social Psychology, 32*, 951–955.

Mele, A. (1997). Real self-deception. *Behavioral and Brain Sciences, 20*, 91–102.

Miller, D. T., & Ross, M. (1975). Self-serving biases in causal attribution: Fact or fiction? *Psychological Bulletin, 82*, 213–225.

Murray-Smith, R., Ramsay, A., Garrod, S., Jackson, M., & Musizza, B. (2007). Gait alignment in mobile phone conversations. In *Proceedings of the 9th international conference on human computer interaction with mobile devices and services. ACM international conference proceeding series* (Vol. 309, pp. 214–221).

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Science, 12*, 367–373.

Paulhus, D. L. (2008). Self deception. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of social psychology* (pp. 189–190). New York: Sage Press.

Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.

Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and the voter's illusion. *Journal of Personality and Social Psychology, 46*, 237–248.

Sloman, S. A. (2005). *Causal models: How we think about the world and its alternatives*. New York: Oxford University Press.

Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences, 10*, 407–412.

Sloman, S. A., & Lagnado, D. (2005). Do we "do"? *Cognitive Science, 29*, 5–39.

Spanos, N. P. (1982). Hypnotic behavior: A cognitive, social, and psychological perspective. *Research Communications in Psychology, Psychiatry, and Behavior, 7*, 199–213.

Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*. New York: Springer.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 216–227.

Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.

Wegner, D. M., & Wheatley, T. P. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist, 54*, 480–492.

Weiner, B. (1992). *Human motivation: Metaphors, theories, and research*. Newbury Park, CA: Sage.