

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

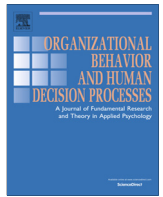
In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/authorsrights>



Contents lists available at ScienceDirect

Organizational Behavior and Human Decision Processes

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

Effort denial in self-deception

Philip M. Fernbach^{a,*}, York Hagmayer^b, Steven A. Sloman^c^a University of Colorado at Boulder, Leeds School of Business, United States^b University of Goettingen, Institute of Psychology, Germany^c Brown University, Department of Cognitive, Linguistic and Psychological Sciences, United States

ARTICLE INFO

Article history:

Received 6 June 2012

Accepted 31 October 2013

Available online 23 November 2013

Accepted by Madan Pillutla

Keywords:

Self-deception

Motivated reasoning

Causal reasoning

Intervention

Causal model

Effort denial

Diagnostic reasoning

ABSTRACT

We propose a mixed belief model of self-deception. According to the theory, people distribute belief over two possible causal paths to an action, one where the action is freely chosen and one where it is due to factors outside of conscious control. Self-deceivers take advantage of uncertainty about the influence of each path on their behavior, and shift weight between them in a self-serving way. This allows them to change their behavior to provide positive evidence and deny doing so, enabling diagnostic inference to a desired trait. In Experiment 1, women changed their pain tolerance to provide positive evidence about the future quality of their skin, but judgments of effort claimed the opposite. This “effort denial” suggests that participants’ mental representation of their behavior was dissociated from their actual behavior, facilitating self-deception. Experiment 2 replicated the pattern in a hidden picture task where search performance was purportedly linked to self-control.

© 2013 Elsevier Inc. All rights reserved.

Introduction

In a classic demonstration of self-deception participants were told that pain tolerance was indicative of the quality of their hearts and then were asked to endure a painful stimulus. In one condition, high tolerance was purportedly indicative of a good heart whereas in another condition, the opposite was true. Those told that high tolerance indicated a good heart endured the pain longer on average than those told the opposite, suggesting that some participants modulated their tolerance to create positive evidence. Moreover, participants denied any influence of the cover story on their behavior. They may even have become more confident about the quality of their hearts after enduring the pain (Quattrone & Tversky, 1984; for similar demonstrations see Chance, Norton, Gino, & Ariely, 2011, and Sloman, Fernbach, & Hagmayer, 2010).

We refer to such behavior as diagnostic self-deception because it involves drawing an invalid diagnostic inference in favor of a desirable trait. This phenomenon violates the logic of causality in that an action expressly taken to support belief in a desirable attribute provides no such evidence. In causal terms, changing one's behavior is an intervention that invalidates the diagnostic relation between behavior and its normal causes (Hagmayer & Sloman, 2009; Meek & Glymour, 1994; Pearl, 2000; Sloman & Hagmayer,

2006). Heart type cannot be responsible for pain tolerance to the degree a participant has manipulated his or her tolerance in response to a cover story.

Diagnostic self-deception is a puzzle for theories of cognition. To draw the self-serving inference, one must remain unaware that one is acting to generate desirable evidence. Yet, in order to generate the evidence, one must perform a causal analysis to determine the desired outcome before executing the behavior. How can one engage in such planning and action and yet remain unaware of doing so when subsequently drawing the beneficial inference?

We propose a solution to this puzzle that builds upon ideas proposed by Quattrone and Tversky (1984) and Sloman et al. (2010). Both of these papers argue that self-deception entails a contradiction between an action and the agent's mental representation of that action. For instance Quattrone and Tversky write that “people select actions to infer a [...] cause, then, to accept the inference as valid, they often render themselves unaware of their having selected the action just to infer the cause (p. 239)”. They describe this as a substitution of a diagnostic for a causal contingency. According to this account, the true contingency is causal (participants in their study chose their pain tolerance based on the cover story) but people treat it as diagnostic (participants attributed their tolerance to their heart type).

Sloman et al. (2010) unpacked this idea and gave it a more precise meaning in terms of causal models. According to their interpretation, there are two paradigmatic ways to represent a choice

* Corresponding author. Address: University of Colorado at Boulder, Leeds School of Business, 419 UCB, Boulder, CO 80309-0419, United States.

E-mail address: philip.fernbach@gmail.com (P.M. Fernbach).

or action that map onto the two types of evidential relations in Causal Model Theory (Pearl, 2000). These models are depicted in Fig. 1. In the *observational model*, some outside factor that is not under willful control, like a personal trait, skill, physical feature, or preference is the cause of a behavior. This causal relation is represented by the arrow from the underlying causes to the choice/action in Fig. 1. In this model the decision-making process is bypassed and, in that sense, the agent is an observer of his or her own action, just as she might be an observer of someone else's (cf. Bem, 1972). Thus the observation of behavior supports diagnostic inference to the underlying causes.

Conversely, in the *interventional model*, the agent represents behavior as due to agency, presumably mediated by a deliberative decision-making process. This model entails that the behavior can be willfully manipulated, and it also entails that the behavior is rendered independent and therefore non-diagnostic of other underlying causes. Hence the interventional model negates the diagnostic relationship between the behavior and these other causes. This is depicted in the figure by the absence of an arrow from the causes not under willful control to the choice/action. Slovic et al. (2010) argued that diagnostic self-deception occurs when people exert some control over their behavior and thus should believe in the interventional model but instead adopt the observational one.

These ideas suggest that self-deception is enabled by adopting the wrong causal model of one's behavior but they do not explain why people would do so. People are generally good causal reasoners and draw reasonable inferences from interventions and observations (Hagmayer, Slovic, Lagnado, & Waldmann, 2007). Moreover, how can people adopt an incorrect causal model of their behavior without becoming aware of the discrepancies between their beliefs and the observable evidence?

We suggest that self-deception is enabled by the inherent uncertainty in the causes of behavior. When self-deceiving, people are clearly manipulating their behavior in a self-serving way to some degree, but this does not imply that their behavior is entirely determined by their will. Other factors (e.g. tolerance for pain) must be influencing their actions too. So both pathways are operative to some extent and there is uncertainty about the contribution of each. It is often impossible to identify with confidence the degree to which a given behavior was freely chosen rather than caused by environmental pressures, personal characteristics, unconscious motives, or bodily states. All we know is what action we have taken and our subjective feeling of choosing. These are not sufficient to rule out either hypothesis. Indeed, people sometimes believe their behavior to be chosen freely and under their personal control even when it is not (Wegner, 2002).

As a consequence, we propose that people have a belief distribution over the two paths, representing assumptions about their influence on behavior. This *mixed belief model* is depicted in Fig. 2. The arrows from the two types of causes to behavior/action

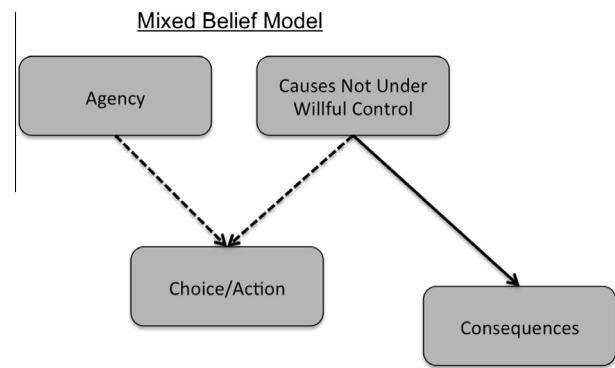


Fig. 2. The mixed belief model of self-deception. Participants distribute belief over the two possible causal paths to action and set the weights on these paths to enable a self-serving inference.

are dashed to indicate that they are malleable and trade off against one another. The implication of this tradeoff is that the diagnosticity of the action for the underlying causes depends on the beliefs about the agency path. Diagnostic self-deception emerges when the underlying causes are associated with good or bad consequences. This creates a motivation to shift beliefs about the two causal paths in a way that increases or decreases the likelihood of consequences depending on whether they are good or bad.

To make this more concrete, consider an example based on the cover story used in Experiment 1. Participants were told that a test provides evidence about how the quality of one's skin will change with age. The test entails enduring a painful stimulus for as long as one can bear. In one condition participants were told that higher pain tolerance indicates lower levels of a chemical in the skin. When present in large quantities this chemical leads to poor skin later in life. In this example, the bad consequences of the chemical induce a motivation to believe that one has low levels of it.

What would bolster such a belief? High tolerance on the test would be necessary but not sufficient to infer low levels of the chemical. The higher the tolerance, the greater the likelihood of low levels of the chemical, but this relation could be explained away by high effort, i.e., by a large contribution of will. Thus in order to draw the diagnostic inference, a self-deceiving participant would have to not only display high tolerance, but also deny exerting great effort to enable the diagnostic inference. Moreover the relation between the belief about the causal contribution of effort and the strength of the diagnostic inference is graded, not all or nothing. The less effort expended to achieve a certain tolerance the stronger the beneficial diagnosis. The mixed belief model therefore predicts that these opposing motivational forces – to display high tolerance by exerting effort and to believe in low effort to enhance the diagnostic inference – will lead to a dissociation between behavior and beliefs about behavior, the hallmark of

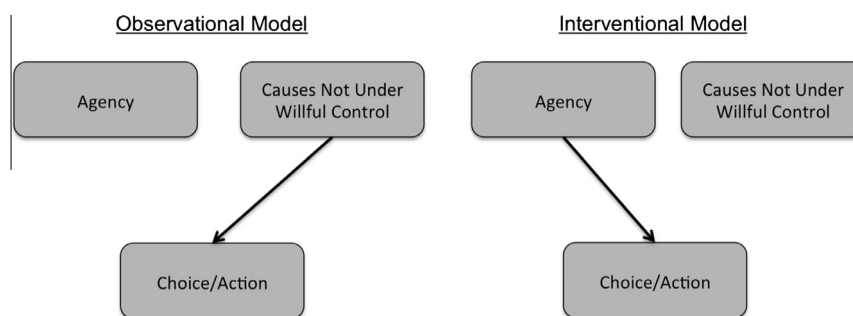


Fig. 1. Two paradigmatic choice models that people can use to construe behavior. In the observational model behavior is treated as due to underlying factors as opposed to agency and thus provides diagnostic evidence. In the interventional model, behavior is non-diagnostic of underlying causes because behavior is represented as freely chosen.

diagnostic self-deception. In this case, participants should exert high effort to display high tolerance but claim low effort to enhance the diagnosticity of performance. We refer to this dissociation between actual and stated effort as “effort denial.” An analogous but opposite prediction would follow if low rather than high tolerance indicated the beneficial diagnosis. In that case the mixed belief model predicts that people will exert low effort to demonstrate low tolerance but will claim high effort to enhance the diagnostic implication of their performance.

Shifts in beliefs about the causal contributions of agency and uncontrolled causes should not be driven entirely by motivation; they should also be constrained by the observable evidence. Participants typically do have some signal, albeit noisy, about the relative contributions of the interventional and observational paths. It would be difficult for a participant in the example above to put in an extreme effort to withstand the pain but maintain the belief that he or she hardly tried. The intensity of the experienced pain would invalidate this belief. [Sloman et al. \(2010\)](#) showed that decreasing uncertainty by providing precise feedback attenuated self-deception, presumably because it made it difficult for people to misrepresent performance as not being modulated by effort when it clearly was. Successful self-deception entails staying within bounds determined by how much uncertainty there is.

Experiment 1: Skin quality and pain tolerance

Women were told that pain tolerance was indicative of the presence of a chemical that determines skin quality later in life. They were either told that high endurance to pain is indicative of good skin in the future (the “high endurance” condition) or the opposite (the “low endurance” condition). The structure of the task was analogous to those implemented by [Quattrone and Tversky \(1984\)](#) and [Sloman et al. \(2010\)](#). There is an unobserved underlying cause (a chemical in the skin) responsible for some desirable attribute (future skin quality). Performance on the task can provide diagnostic evidence about this underlying cause but it cannot influence whether the cause is present. The ability to draw a reasonable diagnostic inference from behavior is contingent on performance being independent of the knowledge about the link between pain tolerance and future skin quality. We expected however that participants would self-deceive by changing behavior to support the beneficial diagnosis.

We further predicted based on the mixed belief model that participants would engage in effort denial to enhance the diagnostic implications of performance. Thus we predicted participants in the high endurance condition would put in a high level of effort relative to those in the low endurance condition but that reported effort would follow the opposite pattern, with those in the low endurance condition reporting high effort relative to those in the high endurance condition.

A second goal of this study is to test how self-deception and effort denial are influenced by explicit feedback about performance. After an initial trial of pain endurance all participants received fabricated feedback that they had endured the pain for a relatively short amount of time. In the low endurance condition this means they are likely to have good quality skin while in the high endurance condition it suggests they will have poor skin. We predicted that this feedback would increase self-deception and effort denial, particularly among those in the high endurance condition.

Methods

Participants and design

Forty-four female students of the University of Göttingen volunteered and were randomly assigned to either the “low

endurance” or the “high endurance” condition. Five participants did not complete the experiment because they indicated that they either perceived very little pain, exerted only low effort to endure the pain, or both. One additional participant had to be excluded because of a data recording error. The remaining 38 participants split equally between conditions.

Materials and procedure

Participants were tested individually in a lab of the Department of Clinical Psychology, and the experiment was allegedly part of a collaboration with the Department of Dermatology. Participants were given detailed but fabricated information about the physiology of the skin and its relation to pain sensitivity and future appearance. In the “low endurance” condition participants were told that Keratohyalin is a naturally occurring chemical that increases the thickness of the skin making it less sensitive to pain, and also causes skin cells to die. In young people, Keratohyalin breaks up before causing visible effects, but starting around age 30 negative effects like rough and porous skin become visible (all participants were well below this age). Hence, skin that is insensitive to pain is likely to have high levels of Keratohyalin and will therefore be of poor quality. In the “high endurance” condition participants were told that Keratohyalin changes the anatomy of the skin such that pain stimuli get transmitted to nerve endings more easily, making skin more sensitive to pain. As in the other condition, they were told high Keratohyalin levels leads to poor skin after age 30. Therefore, skin that is more sensitive to pain is likely to be of worse quality in the future. To increase the credibility of the cover story, participants completed three pages of questions about their skin quality and sensitivity to pain, and some skin cells were rubbed from their hand and stored in a bag.

Next participants were told that a pain stimulus would be applied to their middle finger. After giving informed consent the pain stimulus was applied using an algometer, an apparatus that lowers a spindle onto the middle phalanges of fingers, exerting a pressure of 1710 MPa, and records the duration of the pain stimulation ([Brennum, Kjeldsen, Jensen, & Staehelin Jensen, 1989](#)). By pressing a button participants could lift the spindle to end the pain immediately. All participants were asked to endure the pain for as long as they could, and were not told how long they had endured the pain.

While recovering from the first trial, participants were asked to rate their pain experience on a scale from 1 (rather low) to 5 (unbearable), to describe the type of pain (e.g., pulsating, burning) and to rate their effort on a scale from 1 (hardly any) to 5 (very high). Afterwards all participants were told that they had shown low endurance (in the 7th percentile). Participants then completed the second pain trial, and were asked again to rate their pain experience, describe their pain, and rate their effort. Finally they were asked a series of questions to gauge their credulousness and whether they intentionally changed their endurance due to the cover story.

Results

No participants reported changing their tolerance in response to the cover story and none indicated suspicion of the cover story. We assessed whether participants altered their behavior in the direction that would support a beneficial diagnosis. Because endurance times were not distributed normally we used two-tailed Mann–Whitney *U*-tests to compare these times across conditions. Results are shown in [Fig. 3](#) (standard errors are not shown due to the non-normality of the data). Looking at the pain trials individually, in both cases the low endurance group endured the pain for less time than the high endurance group, but the difference was only

significant in the second trial, $U = 111$, $p = .043$, and was marginally significant overall, $U = 116$, $p = .06$. The difference was not significant in the first pain trial, $U = 128$, $p = .12$.

Our critical prediction was that participants would demonstrate effort denial. After the first trial and before the feedback, participants in the low endurance condition claimed to have exerted significantly greater effort, Mean of 3.6 versus 3.0 (5 was maximal effort); $t(37) = 2.3$, $p = .027$. After the second trial participants in the low endurance group again reported more effort, though the difference between groups was not significant, mean of 3.8 versus 3.6, $t(37) = 1.19$, *ns*. Averaging over both trials, differences in effort were marginally significant, $p = .1$.

We also assessed the correlation between endurance and reported effort. This correlation was negative in both trials, $r = -.33$, $p < .05$ and $r = -.55$, $p < .001$. This implies that participants that changed their behavior the most also engaged in the most effort denial. This could be because these participants were the most motivated to make a beneficial diagnosis about the quality of their skin. This is consistent with previous research showing that there are individual differences in tendency to self-deceive (Starek & Keating, 1991).

Discussion

Women changed their pain endurance to support a beneficial diagnosis of their future skin quality, but reported effort was in the opposite direction of this change. There were some differences between the first and second trial in the pattern of self-deception. The difference in endurance across conditions was only significant in the second trial. This could reflect extra motivation induced by the feedback as we predicted. In fact, there was a differential shift in performance across conditions from the first to the second trials. In the low endurance condition 14 of 19 participants reduced their endurance on the second trial, while only 5 of 19 did so in the high endurance condition (Fisher's exact test, $p < .01$). Effort judgments only differed significantly across conditions in the first trial,

however in both trials the correlation between endurance and stated effort was negative. In sum, the results confirmed our predictions of self-deceptive behavior change and effort denial, however the data were not perfectly consistent. The goal of Experiment 2 was to replicate the results in a different paradigm.

Experiment 2: Searching for hidden pictures

In this experiment we asked participants to find hidden objects in a visual display. They were told that the number of objects people find is indicative of the type of neurally-based visual search they have. People with 'holistic' search take in entire scenes, which makes it hard to notice details and as a result they tend to find few hidden objects. In contrast, people with 'detailed' search focus more narrowly, which causes the hidden objects to pop out, making the task easier. As in Study 1, the underlying trait was linked to a valued effect, in this case self-control and associated behaviors. Thus performance on the search task was purportedly indicative of future health outcomes.

We also simplified the method by omitting the feedback and we created a new procedure to increase self-deception. Instead of presenting all participants with information regarding the relation of visual search and self-control at the outset of the experiment, we first asked them sensitive questions about their body type and eating behaviors and gave them the option to receive more information about why we asked these questions. Only participants who opted into receive additional information were exposed to the cover story manipulation. This had two benefits: It made the cover story more credible and it increased sampling of engaged participants who were interested in the diagnostic value of their visual search behavior.

Finally, we added additional questions to see what inferences participants drew from their motivated behavior change. Self-deception entails not only changing one's behavior by intervening to obtain the desired performance, but also drawing a diagnostic inference from that behavior to the beneficial underlying causes. Again we expected participants to show self-deceptive behavior by manipulating the amount of effort they spent on the task and to deny doing so. In addition, we expected them to draw self-serving diagnostic inferences from the behavior despite having manipulated their performance.

Methods

Participants and design

Three-hundred-fifty-four residents of the United States (52% female, Mean Age = 32.0) were recruited through Amazon Mechanical Turk and were paid 50 cents to complete the study. One-hundred-thirty-six of these participants opted into learn more information and therefore received the cover story about the relation between visual search and self-control. Of these, 29 (21%) failed an instruction check at the end of the experiment and were not included in the analysis. This left a total sample of 107 participants who were assigned at random either to the detailed or holistic condition. In the detailed condition, detailed search (and therefore good performance on the puzzle) was purportedly linked to self-control and beneficial health outcomes. Conversely, in the holistic condition, holistic search was purportedly linked to self-control and beneficial health outcomes.

Procedure and materials

Participants were first instructed that they would search for objects hidden in a picture. They then learned about the two types of visual search, holistic and detailed (see Appendix for materials). After this participants were told that they would be asked to

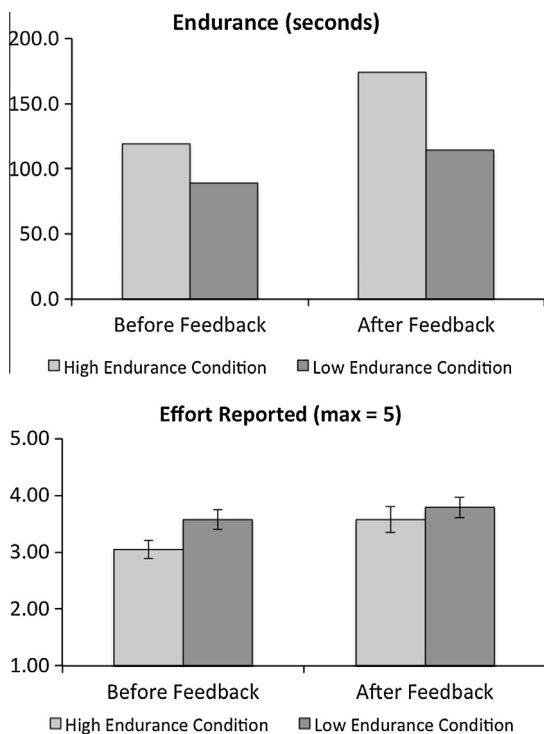


Fig. 3. Results of Experiment 1; Top: Average endurance in seconds by condition before and after feedback (standard errors are not shown due to non-normality); Bottom: Average effort reported by condition before and after feedback with standard errors.

answer a series of sensitive questions and were given the opportunity to exit the survey. Participants that continued were told that they could get additional information about why they were being asked these questions after they answered, and they were asked to indicate whether they wanted additional information or not. Next they answered three questions about body type and eating behaviors (see Appendix for question wordings). If a participant opted into learn more about why we were asking these questions, he or she was told that there was a link between visual search and self-control and that performance on the search task could therefore provide evidence about the likelihood of experiencing beneficial health outcomes. Thus only participants who opted in received the critical information that would create an incentive to self-deceive.

For those participants who opted in, the information about the link between visual search and self-control differed across conditions. In the detailed condition participants were told that detailed search is linked to self-control because “It is believed that the same neural systems that allow people to focus deeply on the details of a scene allow people to retain self-control and avoid urges.” In contrast, in the holistic condition they were told that *holistic search* is linked to self-control because “It is believed that the same neural systems that allow people to focus globally and drown out visual details of a scene allow people to ignore urges and therefore exert self-control.” In both conditions participants were told that “search is therefore connected to positive health outcomes, including less susceptibility to obesity, binge eating and type II diabetes. . . . Moreover, since search type is substantially genetically determined, these effects run in families.” (see Appendix for complete materials). Participants in the detailed condition were therefore motivated to perform well on the task whereas the opposite was true for participants in the holistic condition. This is analogous to the high endurance and low endurance groups in Experiment 1.

All participants then proceeded to a screen where they were taught about the puzzle task. The puzzle was a cartoon drawing of an everyday scene in which objects are hidden (see Fig. 4). Prior to completing the search task, they were instructed about the difference between hidden objects and those that would not count as hidden. They were shown the puzzle with three hidden items circled in green (the bell, pine tree and paper clip) and two non-hidden items circled in red (the mailbox and the bird). Participants were told that they would be asked to find as many objects as they could, but not to start searching yet. They were also told that they should not include any hidden objects that were already pointed out.

Next they were asked whether they had seen the picture before and to indicate the type of search they thought they had on a 7-point scale anchored at “definitely holistic” and “definitely detailed.” After answering these questions they proceeded to a new screen with the same puzzle, which now did not have the green or red circles. Below the puzzle image was a text box and participants were instructed to write down the names of the hidden objects they found. Beneath this was a pull-down menu where they were instructed to enter the total number of hidden objects they found. Participants could spend as much time as they liked on the puzzle. When they decided to proceed, they were asked to rate their level of effort on a seven-point scale and to again indicate what kind of search they thought they had, on the same seven-point scale as before.

After completing these questions participants answered demographic questions and an open-ended instructional check to verify they had understood the cover story (see Appendix). Finally, participants were debriefed about the deception.

Results

We evaluated motivated behavior change by looking at the number of hidden objects that participants found and the amount of time spent on the puzzles across conditions. Results are shown



Fig. 4. The picture search puzzle from Experiment 2.

in Fig. 5. As predicted, participants changed their behavior on the task in line with a beneficial diagnosis. Those in the detailed condition found more objects, condition, $M_{\text{detailed}} = 6.14$, $M_{\text{holistic}} = 5.09$, $t(105) = 2.46$, $p = .016$. Participants in the detailed condition also spent longer on the puzzle, 204.4 s versus 151.5 for the holistic condition, $t(105) = 2.88$, $p = .005$.¹ This replicates previous demonstrations of motivated behavior change (including Experiment 1) in a novel task.

Next we tested the critical novel prediction, that participants would demonstrate effort denial. Confirming the prediction, effort judgments differed across conditions in the opposite direction from behavior change (Fig. 5, bottom panel). Participants in the detailed condition reported less effort, mean of 5.5 versus 6.0 for the holistic condition (7 is maximal), $t(105) = -2.18$, $p = .032$. These results show that participants displayed effort denial; those in the detailed condition tried harder but reported putting in less effort compared to those in the holistic condition.

We also computed correlations between our measures of actual effort (time spent and items found) and reported effort. In this study there was no correlation between time spent and stated effort, $r = .04$, ns . There was a marginally significant negative correlation between number of objects found and stated effort, $r = -.16$, $p = .1$. This provides some evidence that more motivated participants who found more objects also engaged in more effort denial. It is unclear why this negative correlation arose only for objects found and not time spent. In both cases participants who spent more effort did not concede doing so, resulting in no or a slightly negative correlation.

Finally we assessed whether participants ended up with different beliefs about their type of search across conditions. We asked participants twice to report the likelihood they have holistic versus detailed search, once before the puzzle and once after the puzzle. The difference between these two judgments reflects the effect of puzzle performance on belief. Since higher numbers reflect greater belief in detailed search, we expected that participants in the

¹ Reaction times were log normalized for this test. The result is the same with raw reaction times.

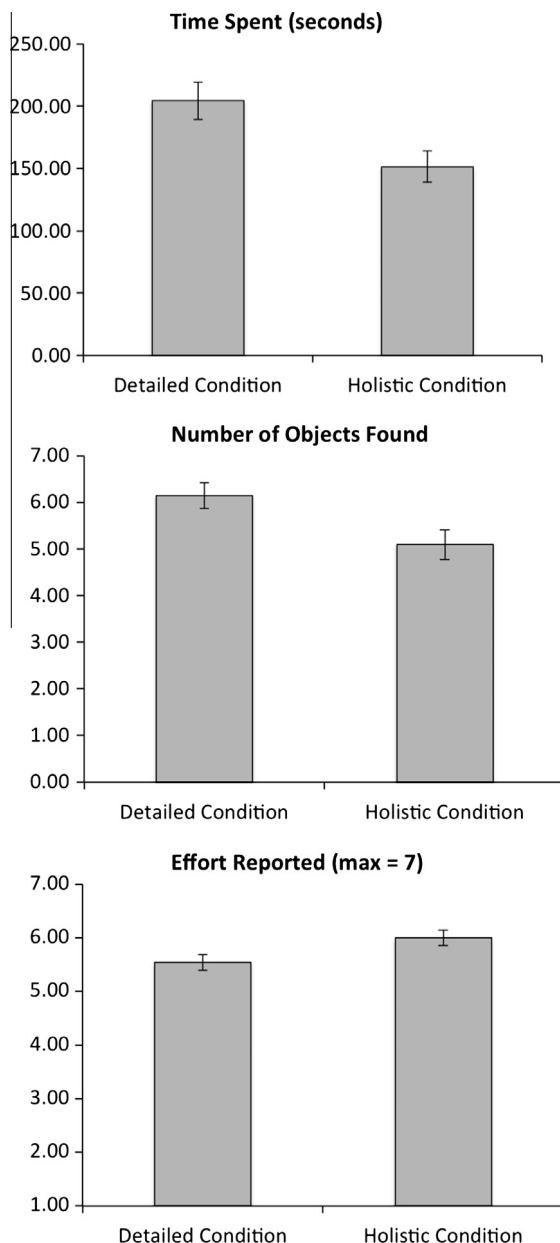


Fig. 5. Results of Experiment 2: Top: Average time spent in seconds by condition with standard errors; Middle: Average number of objects found by condition with standard errors; bottom: Average effort reported by condition with standard errors.

detailed condition would have a higher difference score than those in the holistic condition. This was indeed the case, $M_{\text{detailed}} = .06$, $M_{\text{holistic}} = -.43$, $t(105) = 2.08$, $p = .040$.²

Discussion

Participants in the detailed condition put in greater effort than those in the holistic condition, indicated by more time

spent on the puzzle and more objects found. Ironically, participants engaged in effort denial and therefore those in the holistic condition reported greater effort than those in the detailed condition. This finding supports the mixed belief model. Those in the detailed condition claimed relatively low effort, enhancing the diagnosticity of their performance for detailed search, and hence for self-control. Likewise, those in the holistic condition claimed relatively more effort, thereby increasing the likelihood of holistic search given performance. Moreover, participants actually made the inferences enabled by their differing construals as indicated by the difference in change in beliefs about search type across conditions.

General discussion

We have demonstrated that people alter their behavior to provide evidence that they have a positive trait, yet the effort they report does not reflect the effort they put in. This ‘effort denial’ suggests that self-deception is enabled by people’s tendency to adopt a mental representation of their own behavior that yields the most beneficial inference. As people change their behavior to provide positive evidence for a desirable trait, they simultaneously deny doing so in order to enhance the diagnosticity of the evidence for the positive trait.

Defining self-deception

Since the pioneering work of Gur and Sackeim (1979) self-deception has commonly been defined as entailing simultaneous, contradictory beliefs. Mele (1997) challenged all evidence of self-deception in the psychological literature on the grounds that no empirical demonstrations unambiguously satisfy this definition. We agree that no clear evidence for simultaneous contradictory belief has been found, and we take this as further motivation for the model we propose. We count our results as demonstrations of self-deception not because they involve holding contradictory beliefs. The belief in the agency path does not strictly contradict the belief in the path between uncontrolled causes and action. These beliefs, however, do trade off against each other. The mixed belief model posits that uncertainty in the true causes of action along with the action itself and associated feedback work together to generate self-deception and effort denial. In other words, self-deception does not involve a contradiction among beliefs concerning the causes of behavior, but between a belief (that one is not manipulating behavior) and an action.

Other theories of self-deception

A common way that self-deception has been conceptualized across several disciplines is via analogy to interpersonal deception. According to this account an individual has more than one compartmentalized self. Self-deception occurs when one acts as the deceiver (i.e., the self choosing a behavior) and the other as mark (i.e., the self deriving the beneficial inference). This ‘multiple selves’ idea is discussed at length in philosophy (Pears, 1984; Rorty, 1988), law (Posner, 1997), behavioral economics (Ainslie, 1992), and evolutionary psychology (Von Hippel & Trivers, 2011).

The multiple selves idea also inspired the most rigorous theoretical model of self-deception published to this point, the self-signaling model of Mijovic-Prelec and Prelec (2010; cf. Bodner & Prelec, 2003). According to this theory, behaviors have two types of utility, the utility of the consequences resulting from the behavior and the diagnostic utility of the dispositions that the behavior signals. When making a choice, both types of utilities affect deci-

² As the means show, the difference between conditions is primarily driven by those in the holistic condition increasing belief in holistic processing. Those in the detailed condition show no change in belief about their search type. This probably has to do with the difficulty of the puzzle. Since the puzzle was quite difficult and some of the objects were already pointed out in the training, participants tended to feel like they were not doing that well and therefore were more likely to infer holistic than detailed search. In another (unpublished) experiment we used a different puzzle for training and did not point out any of the objects to participants before search. This easier task led to more objects found, and as a consequence participants were more likely to infer that they had detailed search.

sions. Therefore people sometimes engage in behaviors whose consequences seem to have negative utility (e.g., they lose money), because the behavior signals that they have a valued trait (e.g., being generous). Self-deception arises when diagnostic utility has a measurable influence that participants deny. Mijovic-Prelec and Prelec (2010) model this signaling process as an economic game between two players (i.e., two selves within the same person). Although this model explains self-deception, it does not explain effort denial. It predicts that people will exert more effort to show a behavior that has a positive diagnostic utility than behaviors having none. In causal terms, the model suggests that assumptions about the observational path and the consequences of the uncontrolled causes determine behavior. Agency, as an alternative cause of behavior, is not considered. Therefore this model makes no predictions about participants' judgment of their own effort.

Another theoretical account claims that self-deception is a form of self-serving misattribution due to an under-appreciation of situational influences on behavior. Decision makers are often unaware of how their behavioral choices are guided by context. Self-deception occurs when a decision-maker changes behavior in response to an external cue, but erroneously attributes behavior to an internal disposition (Ariely & Norton, 2008). In self-deception experiments participants change their behavior in a self-serving way while failing to recognize that their behavior is guided by the information provided in the experimental instructions. This account explains why people draw self-serving inferences from observations of their own behavior, but it does not predict effort denial. If people observe themselves, they should recognize how much effort they expended. Our results show this is not the case.

Conclusions

In first demonstrating diagnostic self-deception, Quattrone and Tversky (1984) construed the phenomenon as an error – a cognitive illusion based on people's confusion between diagnostic and causal contingencies. Our results suggest a different conclusion. Far from failing to understand this difference, people appear to be exquisitely sensitive to the diagnostic implications of different causal models; the error is due to motivation, not faulty causal reasoning. Given the complexity of the behavior manipulation and inference processes, diagnostic self-deception might be viewed as a feat rather than a failure of causal reasoning.

We do concede that diagnostic self-deception is an error – at least from an epistemological perspective – in the sense that it leads to a false belief. Even so, the only irrational aspect of people's behavior is that they are giving utility to learning nice things about themselves. If that nice new knowledge inspires beneficial future actions through increased confidence or self-worth, it may be rational in a more global sense (cf. Starek & Keating, 1991).

We attribute the effects we have observed to the inherent uncertainty in knowing the causes of our own behavior. Because we cannot be sure if our actions are governed by our will or by other forces, we take a guess, and that guess is motivated; we construe our behavior in favor of what we want to believe about ourselves. But we are not free to interpret it any way we want. The interpretation is highly constrained by the logic of induction. Vagueness and uncertainty give us a small window to shape interpretations in ways that give us pleasure, hope, and confidence. Thank goodness for small uncertainties.

Acknowledgments

The authors would like to thank Julia Shube, Heidi Jiang, Marina Josephs and Erin Percival for their contributions to the research.

Materials and instructions from Experiment 2

Initial information about puzzle performance and search type

In this study, we are exploring individual differences in the ability to solve visual puzzles. You will look for hidden objects in a picture (the task is similar to children's puzzles that you may be familiar with).

Research on the neuropsychology of vision has shown that there are two types of visual processing that the brain can use to analyze a visual scene.

In **holistic search** the eyes take in the whole scene at once and it is hard to notice details.

In **detailed search** the eyes focus on finer details but are less able to appreciate the whole scene.

Interestingly, people differ in the extent to which holistic or detailed search dominates in their visual processing. Puzzles like the one you are about to do can be used to assess which type someone has. **For people with holistic search, finding the hidden objects will be relatively difficult. For people with detailed search, the objects will seem to 'pop out,' making the task relatively easy**

Questions about body type and self control

How would you describe your weight/body type?

substantially underweight somewhat underweight neither underweight nor overweight somewhat overweight substantially overweight

How often would you say that you eat more than you intend at meals?

Never Rarely Sometimes Most of the Time Always

How often do you eat unhealthy desserts like cookies, cake or ice cream?

Never Very Rarely (once a month) Rarely (once a week) Sometimes (a few times a week) Often (once a day) All the time (multiple times per day)

Cover story information

(Text in bold appeared in the detailed condition. Text in parentheses appeared in the holistic condition)

We understand that these questions may be sensitive for some people to answer. Thank you for completing the survey.

Previous research has suggested that there is a relationship between the type of visual search someone has and their ability to exert self control. **People with detailed search are better at exerting self control, and having detailed search is therefore connected to positive health outcomes, including less susceptibility to obesity, binge eating and type II diabetes. It is believed that the same neural systems that allow people to focus deeply on the details of a scene allow people to retain self-control and avoid urges.** (People with holistic search are better at exerting self control, and having holistic search is therefore connected to positive health outcomes, including less susceptibility to obesity, binge eating and type II diabetes. It is believed that the same neural systems that allow people to focus globally and drown out visual details of a scene allow people to ignore urges and therefore exert self-control.) Moreover, since search type is substantially genetically determined, these effects run in families.

Asking about eating behaviors and body weight allows us to assess this relationship between having detailed visual search and positive health outcomes.

Instructional check (correct answer is "detailed")

To make sure you are reading the instructions carefully, please answer the following questions:

Imagine you are a researcher running this experiment, and you discover the following: Subject C performs in the 90th percentile on the puzzle, meaning he found more hidden objects than 90% of people. Given what you read at the beginning of the experiment what kind of search would you guess the participant probably has?

References

- Ainslie, G. (1992). *Picoeconomics*. Cambridge: Cambridge University Press.
- Ariely, D., & Norton, M. I. (2008). How actions create—not just reveal—preferences. *Trends in Cognitive Sciences*, 12(1), 13–16.
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.
- Bodner, R., & Prelec, D. (2003). The diagnostic value of actions in a self-signaling model. *The Psychology of Economic Decisions*, 1, 105–123.
- Brennum, J., Kjeldsen, M., Jensen, K., & Staehelin Jensen, T. (1989). Measurements of human pressure-pain thresholds on fingers and toes. *Pain*, 38(2), 211–217.
- Chance, Z., Norton, M. I., Gino, F., & Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proceedings of the National Academy of Sciences*, 108(3), 15655–15659.
- 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.
- Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 86–100). Oxford: Oxford University Press.
- Meek, C., & Glymour, C. (1994). Conditioning and intervening. *British Journal for the Philosophy of Science*, 45, 1001–1021.
- Mele, A. (1997). Real self-deception. *Behavioral and Brain Sciences*, 20, 91–102.
- Mijovic-Prelec, D., & Prelec, D. (2010). Self-deception as self-signalling: A model and experimental evidence. *Philosophical Transactions of the Royal Society B*, 365, 227–240.
- Pearl, J. (2000). *Causality*. Cambridge: Cambridge University Press.
- Pears, D. (1984). *Motivated irrationality*. New York: Oxford University Press.
- Posner, R. A. (1997). Are we one self or multiple selves? Implications for law and public policy. *Legal Theory*, 3, 23–35.
- Quattrone, G., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, 46, 237–248.
- Rorty, A. O. (1988). The deceptive self: Liars, layers, and lairs. In B. P. McLaughlin & A. O. Rorty (Eds.), *Perspectives on self-deception* (pp. 11–28). Berkeley: University of California Press.
- Sloman, S. A., Fernbach, P. M., & Hagmayer, Y. (2010). Self deception requires vagueness. *Cognition*, 115(2), 268–281.
- Sloman, S. A., & Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences*, 10, 407–412.
- Starek, J. E., & Keating, C. F. (1991). Self-deception and its relationship to success in competition. *Basic and Applied Social Psychology*, 12(2), 145–155.
- Von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34(1), 1–16.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.