Journal of Experimental Social Psychology 46 (2010) 884–894

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

Inaccurate self-knowledge formation as a result of automatic behavior

Yoav Bar-Anan [a,*], Timothy D. Wilson [b], Ran R. Hassin [c]

[a] Department of Psychology, Ben-Gurion University of the Negev, Beer Sheva, Israel, 84105
[b] University of Virginia, USA
[c] Hebrew University, Israel

ARTICLE INFO

Article history:
Received 30 September 2009
Revised 14 July 2010
Available online 18 July 2010

Keywords:
Self-attribution
Self-knowledge
Goal priming
Automatic social behavior
Confabulation

ABSTRACT

Four studies tested a *post-priming misattribution process* whereby a primed goal automatically influences people's behavior, but because people are unaware of that influence, they misattribute their behavior to some other internal state. People who were primed with a goal were more likely to choose an activity that was relevant to that goal, but did not recognize that the prime had influenced their choices. Instead, people used more accessible and plausible reasons to explain their behavior. The goals were seeking romantic interaction (Studies 1 and 2), helping (Study 3) and earning money (Study 4). People made choices related to these goals but misattributed the choices to temporary preferences (Studies 1 and 3) and more permanent dispositions (Studies 2 and 4). The misattribution had downstream effects, leading to choice behavior consistent with the erroneous self-knowledge. We suggest that automatic behavior can lead to a confabulated self-knowledge with behavioral consequences.

Suppose that Sarah volunteers for a peer tutoring program in her organic chemistry class, in part to satisfy her goal to feel smarter than her fellow students. Suppose further that Sarah is unaware that her desire to feel smarter has affected her choice. How then will she explain her decision to herself? There are likely to be several plausible explanations, such as the possibility that she wanted to help others, meet new people, ingratiate herself with her professor, or add an activity to her application to medical school. Sarah may misattribute her decision to one or more of these alternative goals and not to her goal to feel smarter, resulting in faulty self-knowledge.

The purpose of the present research was to explore this process of misattribution and test hypotheses about its origins, limits, and consequences. We propose a *post-priming misattribution hypothesis* that postulates that a goal or construct can be activated automatically (Step I) and influence people's behavior without their awareness (Step II), but because people are unaware of the actual cause of their behavior (the activated concept or goal) they misattribute their behavior to an accessible and plausible internal state (e.g., a goal, emotion, personality trait, or preference; Step III). As a final result of this process, people incorporate the confabulated internal state into their self-concept and it affects their subsequent behavior (Step IV). Although there is empirical support for each of these steps independently, no prior investigation has looked at the entire sequence of events, from the priming of a goal to confabulated self-knowledge. We conducted four studies that examined the process from beginning to end.

Step I: Internal states can be activated automatically and influence people's behavior

There is considerable support for the first part of our proposed sequence of events, namely that traits, concepts, affect, and goals can be primed in subtle ways that influence interpersonal behavior (Bargh, Chen & Burrows, 1996), judgment (Bargh & Pietromonaco, 1982; Higgins, Rholes, & Jones, 1977) and goal pursuit (Bargh, Gollwitzer, Lee-Chai, Barndollar & Trotschel, 2001; Shah & Kruglanski, 2003). Most relevant to the present work, research has shown that goals can be activated and induce goal-relevant behavior without people's awareness (Aarts, Gollwitzer, & Hassin, 2004; Bargh et al., 2001; Hassin, Bargh, & Zimerman, 2009). In a typical study in this area (Bargh et al., 2001), participants solved a word-search puzzle that included a few words related to cooperation (e.g., helpful, support). Then, as part of what they believed was an unrelated study, participants played a repeated common resources game in which they took the role of a fisherman who could either choose a cooperative strategy (return fish to the lake, so the fish could multiply and help all fishermen) or a competitive strategy (keep the fish). Compared to participants in a control condition, those who received the cooperation words were more likely to share their resources (the fish) with the other fishermen, but were unaware that the word-search puzzle had anything to do with their behavior.

This finding has been replicated using a variety of priming methods (e.g., subliminally presented words, scrambled sentence

* Corresponding author.
E-mail address: baranany@bgu.ac.il (Y. Bar-Anan).

0022-1031/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jesp.2010.07.007

tasks, word-search puzzles, paragraphs that describe someone else's behavior), that activated a variety of goals (e.g., affiliation, impression formation, cooperation, earning money, achievement) that influenced a range of behaviors (e.g., trying to win a ticket to a party, clustering information, sharing resources with others, competing for monetary prizes, learning; Aarts, Custers & Holland, 2007; Chartrand & Bargh, 1996; Bargh et al., 2001; Aarts, Gollwitzer & Hassin, 2004; Eitam, Hassin & Schul, 2008).

Step II: People are unaware of the effects of the primed states on their behavior

Primed participants generally do not attribute their behavior to the priming manipulation (e.g., Fishbach & Labroo, 2007; Sheeran, Webb & Gollwitzer, 2005; Shariff & Norenzayan, 2007). Nor do primed participants report a stronger desire to attain the primed goal than non-primed participants (e.g., Aarts, Gollwitzer & Hassin, 2004; Holland, Hendriks & Aarts, 2005; Fitzsimons & Bargh, 2003), or report pursuing the goal more than do control participants (e.g., Bargh et al., 2001; Chartrand, Dalton & Fitzsimons, 2007; Hassin, Bargh, & Zimerman, 2008). These findings are consistent with the second stage of our proposed sequence of events, namely that people are generally unaware of the nature (and extent) of the influence of the priming manipulations on their behavior. This conclusion, we should note, is consistent with Nisbett and Wilson's (1977) findings that people often make inaccurate reports about influences on their preferences and judgments.

Step III: People misattribute their primed behavior to another internal state

A question that has not been addressed in the priming literature is how people explain their post-priming behavior to themselves. Research on automaticity generally stops at the awareness check, without exploring the downstream effects of priming on self-attribution. We suggest that without awareness of the automatically-activated construct that caused behavior, people often search for other internal states to explain their behavior, thereby forming inaccurate self-attributions.

There is considerable support for the idea that people sometimes attribute their actions to the wrong causes (Bem, 1972; Gazzaniga & LeDoux, 1978; Gazzaniga, 1985; Nisbett & Valins, 1972; Olson, 1990; Ramachandran, 1996). According to self-perception theory, when people's internal states are "weak, ambiguous, or uninterpretable" (Bem, 1972, p. 2), they infer their attitudes and dispositions just as an outside observer would—by observing their behavior and making inferences about why they did what they did. Studies of misattribution have focused on two particular kinds of self-perception errors. In the first paradigm, people are induced to act in an atypical fashion (that is, to do something they would not ordinarily do on their own), but misattribute their actions to a preexisting attitude, trait, or goal. One example of this approach is the induced compliance paradigm from cognitive dissonance studies, in which an experimenter subtly twists people's arms to lie or express beliefs contrary to their attitudes (e.g., Festinger & Carlsmith, 1959). Participants fail to recognize the extent to which their behavior was situationally caused, and mistakenly attribute it to a prior attitude. Such attitude change processes can be fueled by motivational concerns, such as the need to reduce dissonance (Zanna & Cooper, 1974), or can also be the result of a straightforward self-perception process, whereby people misattribute an external cause of their behavior to an internal cause (e.g., Fazio, Zanna & Cooper, 1977; Kiesler, Nisbett, & Zanna, 1969).

A second misattribution paradigm has shown that people can apply the wrong label to internal, physiological cues. Beginning with the classic Schachter and Singer (1962) studies, researchers induced arousal in participants (e.g., with drugs or physical exercise) and demonstrated that under some conditions people misattributed this arousal to emotional states such as sexual attraction (Dutton & Aron, 1974; Cantor, Zillmann & Bryant, 1975; White, Fishbein & Rutstein, 1981) or distress (Fries & Frey, 1980).

Although these paradigms have established important forms of misattribution, we believe that there is another form that is perhaps more common in everyday life but which has not been investigated empirically: high-level internal states such as goals or other constructs are activated automatically but the behaviors that they cause are then misattributed to another internal state. As with Sarah from the opening example, people might act in order to achieve one goal (e.g., to tutor one's peers to satisfy competitive needs), but misattribute their behavior to another internal state (e.g., the desire to help one's fellow students), because they were unaware that the goal was activated and influenced their behavior. Demonstrating this process, we believe, will be an important step in understanding how people develop inaccurate theories about themselves.

Post-priming misattribution has not been previously tested for at least two reasons. First, the idea that internal states can influence people's actions in ways that they do not recognize would suggest the existence of unconscious influences on behavior, including attitudes or goals of which people are unaware. Bem (1972) considered this possibility, but argued that "such claims can edge dangerously close to metaphysics, and . . . should surely be resisted mightily until all other alternatives, save angels perhaps, have been eliminated" (p. 52). Since that time, research on unconscious influences has flourished, however, and it is no longer controversial to suggest that people are unaware of internal states that influence their behavior (Bargh, 2007; Hassin, Uleman, & Bargh, 2005; Nisbett & Wilson, 1977; Wilson, 2002). Second, it took a few decades for researchers to develop techniques whereby high-level causes of behaviors (such as goals) were activated outside of awareness (Aarts, Gollwitzer, & Hassin, 2004; Bargh et al., 2001; Hassin et al., 2008; for a recent review see Ferguson, Hassin, & Bargh, 2008).

Step IV: Self-misattribution leads to inaccurate self-knowledge

Research has shown that people incorporate misattributed internal states into their self-concept and act consistently with them (e.g., Fazio, Effrein & Falender, 1981; Freedman & Fraser, 1966; Gorassini & Olson, 1995). No studies have shown, however, that such misattribution can occur in a priming paradigm. Nor, we should note, did Nisbett and Wilson's (1977) studies explore the impact of a lack of awareness on self-knowledge; for example, they showed that people were unaware that the order of consumer goods influenced their preferences, but did not examine how, if at all, that lack of awareness influenced people's self-concepts. We used modern priming paradigms to test the hypothesis that a primed concept would influence people's behavior, that people would fail to recognize the effect of the primed goal on their behavior, that they would misattribute their behavior to another internal state, and finally, that this confabulated internal state would be incorporated into their self-concept and influence subsequent behavior.

Overview of the studies

We tested our post-priming misattribution hypothesis in four studies in which we primed a goal (e.g., to affiliate with a member of the opposite sex) and then asked people to choose between two alternatives (e.g., to take part in one of two tutoring sessions). One of the alternatives could advance achievement of the goal (e.g., one tutor was a woman and the other was a man), and we predicted that people primed with the goal would be more likely to choose that alternative (e.g., the opposite-sex tutor). The activities also varied according to decoy attributes that could be plausibly used to explain one's choice; for example, the male tutor taught one topic and the female tutor

taught another (counterbalanced). We predicted that primed participants would not be fully aware of how much the primed goal (e.g., to affiliate with a member of the opposite sex) influenced their choice, and would thus misattribute their choice to the decoy attribute (e.g., the topic taught by the tutor).[1]

Across the studies we primed different goals, namely opposite-sex affiliation (Studies 1 and 2), helping (Study 3), and earning money (Study 4). The decoy attributes that participants were expected to use to explain their choice were preferences for different activities, such as types of word games (Studies 1 and 3), or dispositional tendencies, such as a chronic preference for challenging tasks (Studies 2 and 4). Studies 1–3 tested our basic hypothesis that priming can cause faulty self-knowledge formation via a process that is best described as "overattribution," whereby people underestimate the effect of the primed goal, and overestimate the effect of the decoy internal state. Study 4 tested a purer case of misattribution, whereby people come to believe that the decoy internal state influenced their behavior when in fact the structure of the experiment ensures that this state played no role. Study 4 also tested whether the false self-knowledge, acquired by misattribution, would affect subsequent choice behavior.

Studies 1 and 2: The effect of the goal to affiliate on interest in attention versus health

In two studies, male participants were primed with the goal to affiliate with women and then chose to be tutored by a woman or a man who taught different topics. We predicted that the affiliation-primed participants would be more likely than control participants to choose the female tutor, but would (mis)attribute this choice to their interest in the topic she taught. Study 2 was an exact replication of Study 1, with an extension of the self-knowledge dependent measure from temporary preferences to inferences about one's dispositions. We thus present these two studies together.

Method

Participants

Participants were 62 male students (mean age = 19.03, *SD* = 1.6) in Study 1, and 70 male students (*M* age = 19.4, *SD* = 1.43) in Study 2. They received credit in their undergraduate psychology courses (Study 1) or $5 payment (Study 2). We discarded from the analyses participants who indicated that they were not sexually attracted to women (1 in Study 1, 5 in Study 2), and one participant (Study 1) who was uncooperative and listened to his MP3 player during the study.

Procedure

Participants took part individually on computers in what they believed was a study of communication over the internet. Instructions on the computer informed participants that they would complete a few different experiments developed by different researchers.

Goal priming. The first task was used to prime the goal to affiliate with women. Participants read a short passage used in previous research (Aarts, Gollwitzer, & Hassin, 2004) that described a night at a pub from the perspective of a male protagonist. The two last sentences in the affiliation prime passage were: "At the end of the evening, Elliott walks Nicole home. When they arrive at her home, he asks her, 'May I come in?'" Participants in the control group read the same passage with a different ending: "At the end of the evening people start to dance. Elliott looks at the dance floor from a distance, and thinks, 'Isn't this a nice place to be?'".

Choice options. The next task, entitled "The Tutoring Project," was ostensibly designed to help undergraduate experimenters practice an on-line tutoring session for a future experiment. Participants were told that they would be connected to an available tutor who would teach them a topic for 7 min, after which participants would give the tutor feedback on how to improve his or her performance. The program then appeared to establish an internet connection with two available tutors, Jason and Jessica, one of whom would teach about "attention improvement tips and tricks" and the other about "how to prevent common illnesses." The match between tutor and topic was counterbalanced. Participants were instructed to get ready to choose one of the tutoring sessions as soon as an instant messaging program was launched. After 50 s the program failed to launch and an error message instructed participants to call the experimenter. The experimenter, after a pause, told the participant that he could move on to a general questionnaire about the tutoring project. The purpose of this ruse was to allow participants to decide which tutor they preferred without having to report their choice (at least not until after they rated their interest in the tutoring topics), thereby eliminating self-presentation concerns when reporting the reasons for their choice.

Dependent measures

Interest in topics. Participants rated their interest in several topics that were ostensibly used in various tutoring sessions. Among these were the topics taught by Jessica and Jason, "attention improvement tips and tricks," and "how to prevent common illnesses." The scales ranged from *1 = not interested at all* to *9 = very much*.

Self-reported dispositions. In Study 2 we added two questions about participants' dispositions: "*I'm a person who really cares about my cognitive skills*" and "*I'm a person who really cares about my health*," each rated on scales from *1 = strongly disagree* to *9 = strongly agree*. Participants also rated three filler dispositions.

Choice measure. Participants indicated which tutoring session they had preferred on a scale that ranged from *1 = the first, much more* to *8 = the second, much more*.

Reported reasons for choice of tutor. Participants explained the reasons for their choice of tutor in an open-ended format, then rated the importance of four different considerations: the tutor's name, the session's topic, whether the session was listed first or second, and the tutor's gender, all on scales ranging from *1 = not important at all* to *9 = very important*.

Awareness of goal pursuit. Participants were then asked, "How much do you feel motivated to seek a sexual or romantic relationship right now?" on a scale that ranged from *1 = not at all* to *9 = very much*.

Awareness of the effects of the prime. Finally, participants indicated whether they thought that the priming passage had any effect on their choice of tutoring sessions.

Results

Effects of prime on choice of tutor

We recoded people's preference ratings such that high numbers reflect a preference for the female's tutoring session and analyzed

[1] The studies used methods that were previously used to activate goals that led to unconscious goal pursuit (Aarts et al., 2004; Bargh et al., 2001; Custers & Aarts, 2005). However, because our post-priming misattributions hypothesis refers to behavior that follows construct activation, regardless of whether the behavior constitutes goal pursuit or not, we did not verify that the activated behavior was in fact goal pursuit and not construct-related behaviors that do not constitute goal pursuit (e.g., Bargh, Chen & Burrows, 1996; Dijksterhuis & van Knippenberg, 1998). Hence, while we think that the best description of the priming manipulations is *goal priming*, we refrain from calling the priming effects *goal pursuit*. See further discussion of this issue in the General discussion section.

these ratings with a 2 (Prime: affiliation vs. control) × 2 (Topic: female tutor taught attention vs. health) ANOVA. As predicted, participants in the affiliation prime condition reported more of a preference for the female's session (Study 1: $M = 5.83$, $SD = 1.93$; Study 2: $M = 6.03$, $SD = 2.22$) than did participants in the control group (Study 1: $M = 4.77$, $SD = 2.28$; Study 2: $M = 5.09$, $SD = 2.52$), $F(1, 56) = 4.46$, $p < .05$, $\eta_p^2 = .08$ in Study 1, $F(1, 61) = 4.12$, $p < .05$, $\eta_p^2 = .06$ in Study 2. There was also a main effect of Topic, $F(1, 56) = 15.08$, $p < .001$, $\eta_p^2 = .21$ in Study 1, $F(1, 61) = 33.01$, $p < .0001$, $\eta_p^2 = .35$ in Study 2, reflecting the fact that people in both conditions showed a greater preference for the female's session when she taught attention than when she taught health. (In other words, attention was a more popular topic than health). There was no Prime × Topic interaction, $Fs < 1$.

Overattribution effects on self-knowledge

Reported interest in the topics. We predicted that the participants in the affiliation priming condition would overattribute their choice of tutor to the topic she taught. To test this prediction, we computed the difference between participants' reported interest for the female and male's topic. High numbers thus reflect a preference for the topic taught by the female. As predicted, there was a main effect of Prime on this index, $F(1, 56) = 4.7$, $p < .05$, $\eta_p^2 = .09$ in Study 1, $F(1, 61) = 4.44$, $p < .05$, $\eta_p^2 = .07$ in Study 2. This occurred because participants in the affiliation prime condition reported more of a preference for the female's topic (Study 1: $M = 1.14$, $SD = 2.03$; Study 2: $M = 1.03$, $SD = 2.36$) than unprimed participants did (Study 1: $M = -0.13$, $SD = 2.45$; Study 2: $M = -0.44$, $SD = 2.58$). Across priming conditions, people reported more interest in the attention than the health topic, not significantly in Study 1, $F(1, 56) = 2.62$, $p = .11$, $\eta_p^2 = .04$, and significantly in Study 2, $F(1, 61) = 6.12$, $p < .05$, $\eta_p^2 = .09$. There was no interaction, $Fs < 1$.

Personality dispositions. In Study 2 we assessed participants' ratings of their dispositional interest in the topics taught by the tutors. To test the hypothesis that primed participants would attribute their choice to their dispositions, we subtracted participants' ratings of their interest in the topic taught by the male from their ratings of their dispositional interest in the topic taught by the female. A 2 (Prime: affiliation vs. control) × 2 (Topic: female taught attention vs. health) ANOVA revealed the predicted main effect of Prime, $F(1, 61) = 4.82$, $p < .05$, $\eta_p^2 = .07$, reflecting the fact that participants in the affiliation prime condition reported a stronger disposition toward the topic taught by the female ($M\ diff = 1.00$, $SD = 2.16$) than did control participants ($M = 0.03$, $SD = 1.31$). There was neither a main effect of topic, $F(1, 61) = 1.09$, $p = .30$, $\eta_p^2 = .02$, nor an interaction, $F < 1$.

The effect of priming on self-knowledge change. The results so far support our prediction that priming an affiliation goal would make participants more likely to choose the female tutor, but to misattribute this choice to a preference for the topic the female tutor taught. We further explored whether the primed participants were more likely than the control participants to attribute their choice of tutor to the topic. If that is the case, then primed participants who chose the female's tutoring session should show stronger preference for her topic than control participants who chose the female's session. This is a strong test of our hypothesis, we should note, for two reasons. First, control participants, unprimed with affiliation, probably were more interested in the topic of the tutoring session that they chose. Nonetheless, if primed participants were misattributing their choice to the topic, they might show an even stronger preference for the topic than unprimed participants. Second, this test suffered from low power (smaller sample) because it included only participants who preferred the female's session over the male's session. Finally, the staged computer error prevented us from recording the participants' choice. Instead, we can only use their preference between the two tutoring sessions, reported *after* they have already rated the topics.

Despite these limitations, we found that primed participants who preferred the female's tutoring session over the male's tutoring session did in fact express more preference to her topic over the male's topic (Study 1: $M = 1.82$, $SD = 1.74$; Study 2: $M = 1.70$, $SD = 1.49$) than did control participants who preferred the female's session (Study 1: $M = 1.06$, $SD = 1.86$; Study 2: $M = 1.05$, $SD = 1.61$). But, despite moderate effect size, this difference did not reach significance in either study; Study 1: $t(38) = 1.34$, $p = .18$, $d = .42$; Study 2: $t(40) = 1.36$, $p = .18$, $d = .42$. Similarly, in Study 2, primed participants who preferred the female's session reported stronger dispositional interest in the topic taught by the female ($M = 1.22$, $SD = 1.93$), than control participants who chose the female's session ($M = 0.26$, $SD = 1.10$), $t(35.9) = 2.01$, $p = .05$, $d = .61$ (there were 41 participants in that analysis, but the variances of the two groups were unequal). This pattern of results may suggest that priming participants without their awareness made them more likely to use their choice behavior in order to learn about themselves.

Reported reasons for choice of tutor

In all of the studies in this paper, the main evidence for self-misattribution was the distorted self-knowledge just reported. However, we also measured people's reported reasons for why they chose what they did, to learn more about the accuracy of self-attribution.

Priming. Participants were unaware that the priming affected their choice. Only one participant (in Study 1) reported that the story had affected his choice of tutor, saying that it had stirred his interest in women.

Goal desirability. We also asked participants how motivated they were to seek a sexual or romantic relationship, and found no awareness of the goal activation. There was no significant difference on this measure between affiliation-primed participants (Study 1: $M = 5.38$, $SD = 2.32$; Study 2: $M = 5.78$, $SD = 2.11$) and control participants (Study 1: $M = 5.58$, $SD = 2.26$; Study 2: $M = 5.45$, $SD = 2.49$), $ts < 1$.

Reported reasons. Participants rated the extent to which the tutor's name, the session's topic, whether the session was listed first or second, and the tutor's gender influenced their choice. Participants rated the topic as most important (Study 1: $M = 7.77$, $SD = 1.2$; Study 2: $M = 7.26$, $SD = 1.93$), and the other three possible reasons, the tutor's gender (Study 1: $M = 2.72$, $SD = 2.18$; Study 2: $M = 2.48$, $SD = 2.03$), whether the session was listed first or second (Study 1: $M = 2.08$, $SD = 1.85$; Study 2: $M = 1.52$, $SD = 1.13$), and the tutor's name (Study 1: $M = 2.38$, $SD = 2.07$; Study 2: $M = 2.05$, $SD = 1.75$) as much less influential, $ps < .0001$. Importantly, there was no significant effect of the priming manipulation on the reported influence of any of the four variables. Finally, when asked to report their reasons in an open-ended response format, almost all participants (Study 1: 96%, Study 2: 98%) mentioned only the topic as the reason for their choice.

Discussion

As predicted, men primed with an affiliation goal were more likely to choose a female tutor than men in the control condition, but were unaware of the effects of the prime. This finding replicates prior research on goal priming. Consistent with our misattribution hypothesis, men in the priming condition reported that they liked whatever topic that a female tutor happened to teach more than did men in the control condition, and reported more of a dispositional interest in that topic. The topics undoubtedly affected some people's choice, thus we cannot say that primed participants' reported interest in the topics were completely wrong. The fact that primed

participants reported more of an interest in *whichever* topic the female tutor happened to be teaching, however, suggests that people overattributed their choice to the topic. In Study 3 we sought to extend these findings to a different goal.

Study 3: The effect of a helping goal on preference for word games

Studies 1 and 2 had a surface similarity to a couple of studies reported by Bernstein, Stephenson, Snyder and Wicklund (1983). Those studies showed that when one of two alternative activities in an experiment promised men proximity to an attractive woman, fear of rejection caused the men to choose that alternative only when their choice could have been attributed to another feature of that alternative. Unlike our studies, Bernstein et al. did not induce the motivation to affiliate without participants' awareness, and did not address the issue of self-attribution and self-knowledge. However, their results do suggest that romantic motivation might be perceived as non-flattering or threatening. Perhaps in our studies, as well, participants may not have wanted to admit that romantic or sexual feelings influenced their choice of tutor. Such a possibility is consistent with our theorizing, to the extent that it operated outside of people's awareness and kept them from recognizing that the prime influenced their choice.

Alternatively, participants might have been fully aware of the effects of the prime but were reluctant to say so because of self-presentational concerns. Instead of acknowledging to the researchers that they were thinking about romance or sex, perhaps they reported more rational-sounding reasons, such as an interest in the tutor's topic. The purpose of our staged computer error was to minimize this possibility by convincing participants that they would never reveal their choice. Participants rated their interest in the topics *before* revealing which tutoring session they preferred, thereby minimizing self-presentational concerns. Nonetheless, it is important to show that priming can lead to a distortion of self-knowledge even when the primed goals are socially acceptable.

In Study 3, participants were primed with a socially desirable goal —helping others. They then chose to play one of two different word games, one of which involved helping and the other competing. We predicted that primed participants would be more likely to choose the game in which they could help others, but that they would be unaware of the influence of the prime and would misattribute their choice to other features of the game.

Method

Participants

Participants were 76 students (64 females, M age = 19.3, SD = 2.52) who received credit in their undergraduate psychology courses.

Procedure

Goal priming. Participants took part individually in what they believed were unrelated studies. The first task was a word-search puzzle (a matrix of letters in which people find words) that served as the goal priming manipulation (Bargh et al., 2001). In the priming condition, 6 of the 14 words in the puzzle were related to helping (*donate*, *oblige*, *give*, *support*, *generous*, *kind*) and eight were not (e.g., *chair*). In the control condition, none of the words was related to helping.

Choice. Participants were then asked to choose to play one of two word games: "hangman," in which they would guess names of movies by selecting letters one at a time, and "missing letters," in which they would guess names of animals from incomplete words. The "mode" of each game was also described, namely that one involved helping the researchers (by providing feedback and suggesting new words for the puzzle) whereas the other involved competing ("strive to excel beyond the achievements of past participants"). We counterbalanced whether the opportunity to help was associated with the hangman or missing letters game. If primed participants preferred the game in which they could help, but were unaware of the effects of the prime, they could attribute their choice to either the type of game (hangman vs. missing letters) or the topic (movies vs. animals) or both.

Reasons for choice and interest in topics. After choosing a game, participants were asked to describe the reasons for their choice in an open-ended format. They then rated how interested they were in the topics of the games (animals, movies), the types of games (hangman, missing letters), and in the mode of the games (helping, competing), in random order, on scales ranging from *1 = not at all* to *9 = very much*. Next, participants rated the influence of each attribute of the games (topic, type, mode) on their choice, in random order, on scales ranging from *1 = not important at all*, to *9 = very important*. They also rated the influence of the desire to feel competent, helpful, outperform someone else, or help the researchers.

Awareness of goal activation. Participants were also asked how much they felt like helping others and competing with others during the study, on scales that ranged from *1 = not at all*, to *9 = very much*.

Awareness of the effects of the prime. Participants were probed about whether they noticed a theme in the word-search puzzle (the priming manipulation) and whether they thought that the puzzle affected their choice of game.

Results

Effects of prime on choice of game

We performed a 2 (Prime: help vs. control) × 2 (Topic: opportunity to help paired with hangmen/movies vs. missing letters/animals) factorial logistic regression on participants' choice of game. As predicted, help-primed participants chose the game in which they could help more often than did control participants (61% vs. 38%), χ^2 (df = 1, N = 76) = 3.71, p = .05. There was neither a main effect of which game was paired with the opportunity to help nor an interaction, χ^2s < 1, ps > .61.

Overattribution effects on self-knowledge

We tested the effect of priming on self-reported liking of the various attributes of the two alternatives with a 2 (Prime: help vs. control) × 2 (Topic: opportunity to help paired with hangmen/movies vs. missing letters/animals) ANOVA on three preference scores, one for each attribute dimension (type of game, game topic, and whether the game involved helping or competition).

As predicted, primed participants reported a significantly greater preference for the type of game (hangman vs. missing letters) that happened to involve helping (M difference = 0.70, SD = 2.71) than did control participants (M = −0.42, SD = 2.43), F(1, 72) = 3.83, p = .05, η_p^2 = .05. There was also a significant effect of Game type, F(1, 72) = 11.39, p < .01, η_p^2 = .14, reflecting the fact that, across priming conditions, participants preferred hangman more than missing letters. There was no interaction, F(1, 72) = 1.31, p = .26, η_p^2 = .02.

Primed participants also reported a greater preference for the topic paired with helping (movies or animals, M = 0.70, SD = 2.85) than did control participants (M = −0.19, SD = 2.99), but this difference was not significant, F(1, 72) = 1.67, p = .20, η_p^2 = .02. Neither the main effect of which game was paired with the opportunity to help nor the interaction was significant, Fs < 1.

Finally, we asked participants how interested they were in the opportunity to help and to compete. There was no effect of prime on these measures, Fs < 1, suggesting that primed participants did not notice that this attribute was more attractive to them than the usual. On average,

participants reported more interest in competing ($M = 5.57$, $SD = 2.20$) than helping ($M = 4.46$, $SD = 1.91$), $t(75) = 3.44$, $p < .01$, $d = .54$.

The effect of priming on self-knowledge change. As in Studies 1 and 2, we examined whether primed participants attributed their choice to the decoy feature more than did control participants. Like in Studies 1 and 2, the pattern of results suggested that primed participants who chose the activity that involved helping reported more liking for the type of game that they chose ($M = 1.42$, $SD = 2.96$) than did control participants who chose the activity that involved helping ($M = 0.14$, $SD = 2.88$). Nevertheless, this difference was not significant, $t(36) = 1.29$, $p = .20$, $d = .44$. Again, we note that this is a particularly strong test of that hypothesis because control participants, who were not primed, presumably did prefer one type of game over the other, accounting for their choice.

Reported reasons for choice of game

Priming. None of the participants in the prime condition correctly identified the theme of the word-search puzzle or reported that the puzzle had affected their choice of game.

Goal desirability. Help-primed participants did not rate themselves as having more of a desire to help ($M = 4.44$, $SD = 2.12$) or less of a desire to compete ($M = 3.9$, $SD = 2.43$), than control participants (M help $= 4.33$, $SD = 2.01$, M compete $= 4.17$, $SD = 2.21$), $ts < 1$.

Reported reasons. Participants rated how much their choice of game was influenced by specific attributes of the games and various motives. The priming manipulation did not significantly affect any of these measures, $ts(73) < 1.54$, $ps > .13$. Participants rated the type of game ($M = 6.04$, $SD = 2.26$) and topic of the game ($M = 6.13$, $SD = 2.05$) as most influential, followed by the game's mode (helping vs. competing), ($M = 5.32$, $SD = 2.28$), $t(75) = 2.26$, $p = .03$ for the comparison between the influence of the topic of the game and the influence of the mode. Finally, no participant mentioned in the open-ended responses that he or she chose a game because of the opportunity to help.

Discussion

Participants who were primed with helping were more likely to choose the word game that involved helping, but appeared to be unaware that this was a reason for their choice (despite the fact that helping is a socially desirable goal).[2]

Instead, primed participants overattributed their choice to the type of game. If hangman happened to be paired with the opportunity to help, they reported a preference for hangman, whereas if missing letters happened to be paired with opportunity to help, they reported a preference for missing letters (relative to participants in the no prime control condition).

Study 4: The effect of post-priming misattribution on subsequent choices

Studies 1–3 demonstrated that when goal priming affects people's behavior, they generate inaccurate accounts to explain their behavior and thus acquire inaccurate self-knowledge. We sought to extend these findings in Study 4 in two main ways. First, we examined whether the new, inaccurate, self-attribution would affect people's subsequent choice behavior, in addition to their self-reports about their interests and dispositions. Second, we examined whether unconscious goal activation can lead to the complete confabulation of a reason for one's choice. Studies 1–3 established that unconscious goal activation can lead to overattribution, whereby people exaggerate the extent to which their choices were influenced by choice attributes that might have had some effect (e.g., the topics the tutors taught in Studies 1 and 2). In Study 4 we tested the hypothesis that people will confabulate a reason that could not have had any influence on their choice. We did so by introducing a decoy attribute after participants had made their choice and which thus could not have influenced their decision. We predicted that people might misremember when they learned about the decoy attribute and thus mistakenly attribute their choice to it.

Study 4 also extended the generalizability of the findings by using a different manipulation to prime a different goal. We primed participants with the goal of earning money and then asked them to choose to play one of two trivia games. For the priming manipulation, people read a passage about someone who attempts to earn money. They then saw pictures depicting each trivia game, one of which (counterbalanced) included the pictures of presidents as they appear on United States (U. S.) currency. Even though participants could not earn money from either game, we predicted that priming money would make them more likely to select the game that had money in its picture. As in the previous studies, we expected that participants would not recognize the effects of the prime on their choice and would instead attribute their choice to a decoy attribute of their preferred game, in this case how easy or challenging it was said to be.

Method

Participants

Participants were 127 students (89 females, M age $= 18.72$, $SD = 1.53$) who received credit in their undergraduate psychology courses.

Procedure

Goal priming. Participant took part individually in what they believed were unrelated studies. As in Studies 1 and 2, the goal was activated using the automatic goal contagion manipulation. Participants in the priming condition read a passage describing a fellow student from their university who plans to earn money, whereas participants in the control condition read a similar passage describing the person's plans to return a CD to a friend. Following past research, the character's gender and university affiliation matched the gender and university affiliation of the participant (Loersch, Aarts, Payne & Jefferis, 2008).

Choice. Next, participants chose to play one of two trivia games called "American Government" and "American Politics." The games were portrayed in two color pictures that appeared side-by-side on the computer screen (see Fig. 1 for a black and white version). Each picture showed a collage of images of people and symbols relevant to the trivia topic. In one of the pictures (counterbalanced) we inserted images of U. S. presidents as they appear on $1, $10, and $20 bills. We

[2] We conducted a follow-up survey of participants from the same population to test our assumption that helping was a socially desirable goal ($N = 77$). Participants read a description of the two choice alternatives that were presented in Study 3 and rated how embarrassed they would have been if people had known that the reason for their choice was that they wanted to (1) compete; (2) help; or (3) that they preferred the game in that alternative. They also reported how much they would have been disappointed in themselves, had they chosen one of the alternatives because of these three reasons. The response scale ranged from 1 (*not at all embarrassed* [*disappointed*]) to 7 (*very embarrassed* [*disappointed*]). Helping was rated as significantly less embarrassing than competing ($Ms = 1.43$ vs. 2.05, $SDs = .87$, 1.34), $t(76) = 4.12$, $p < .001$, $d = 0.55$, and as significantly less disappointing to the self ($Ms = 1.42$ vs. 2.12, $SDs = .95$, 1.48), $t(76) = 3.18$, $p = .002$, $d = 0.56$. Helping was also rated as less embarrassing than preferring the game ($M = 1.71$, $SD = 1.23$), and less disappointing than preferring the game ($M = 1.68$, $SD = 1.28$) with marginal significance in both scales, $ts(76) = 1.89$, 181, $ps = .06$, .07, $ds = 0.26$, 0.23, respectively. These results are consistent with our assumption that participants viewed helping as a socially desirable goal; indeed, the mean ratings of 1.43 and 1.42 were close to the endpoints of the scales.

Fig. 1. Black and white version of the cover pictures of the games in Study 4. Participants saw this picture for 13 s before selecting one of the games. The location of the currency images was counterbalanced between participants (either in cover picture of *American Government* or *American Politics*).

reasoned that American presidents are relevant to both topics (American Government and American Politics) and thus could plausibly go with either trivia game. The two pictures appeared together for 13 s, at which point the participants chose the game they wanted to play.

Decoy: difficulty of the of trivia games. Participants then learned that one of the trivia games was "pretty challenging" and the other "fairly easy." We counterbalanced which game received which label, such that half of the participants learned the game they had just chosen was challenging and half learned that it was easy. Because people did not receive this information until after they had made their choice it could not have influenced which game they chose. After a few more slides of instructions about the trivia game, we attempted to obscure participants' memory for this sequence of events by asking them to indicate again which game they chose before (all participants repeated the same choice).

The trivia games. Participants were then told that they would receive questions from both trivia games. There were two sets of eight questions, each comprised of four questions pertaining to each topic (presented in intermixed order). In the set that was presented to participants who had been told that the American Politics game was more difficult, the four questions about American politics were relatively difficult and the four questions about American government were easy. In the set that was presented to participants who had been told that the American Government game was more difficult, the four questions about American government were relatively difficult and the four questions about American politics were easy. Notice that participants' choice of game did not influence what questions were presented to them. Only the random assignment of one of the topics as the difficult topic and the other topic as the easy topic influenced which of the two sets they completed. Therefore, all participants answered four difficult questions and four easy questions. Participants were not told whether their answers were correct.

Dependent measures

Dispositional liking for challenge. On a seemingly unrelated filler questionnaire, participants rated how much they liked challenges on a scale that ranged from *1 = not at all* to *9 = very much*.

Ratings of interest. Participants rated their interest in a few topics, including the two trivia game topics, on a scale that ranged from *1 = not at all* to *9 = very much*.

Choice of "tips". Participants then waited a few more seconds while the computer appeared to choose a new study for them. In the "new" study, participants were told that they would read a list of five tips about one topic and complete a memory test about these tips. They were then asked which list of tips they preferred to read: "How to make and save money" or "How to successfully pursue challenges." We reasoned that if people had misattributed their initial choice of game to a preference for its level of difficulty, this confabulated self-perception should influence how likely they were to chose the tips about pursuing challenges. That is, participants who learned that their choice of game was challenging should be *more* likely to choose the tips about pursuing challenges, whereas participants who learned that their choice of game was easy should be *less* likely to choose tips about pursuing challenges. The fact that the alternative activity (tips about money) was consistent with the primed goal allows an especially strong test of the confabulation hypothesis: will primed participants choose an activity that is relevant to the primed goal (tips about money) or an activity that is consistent with their newly-formed, confabulated self-perception (that they prefer easy or difficult games)? After participants chose the list of tips, they were presented with the tip list.

Reasons. Participants were told that the next few questions were about the study that involved the trivia games. First, they described the reasons for their choice of game in an open-ended format. They then rated the extent to which nine potential reasons influenced their choice of game: difficultly (whether the topic was easy or difficult), topic (whether the topic was interesting), cover picture (whether it was attractive), the fact that money appeared on one cover picture, and their desire to earn money. The other four served as filler reasons: the desire to gain knowledge, avoid risks, relax, and appear attractive. The scales ranged from *1 = not at all*, to *9 = very much*.

Awareness. Participants were asked whether the passage about earning money affected their behavior in the rest of the study, and, specifically, whether it affected their choices. They also reported how much they were motivated to earn money at that moment, on a scale that ranged from *1 = not at all*, to *9 = very much*.

Design. The design thus consisted of 8 between-participant conditions: 2 (Prime: money vs. control) × 2 (Game with Money in Picture: American government vs. American politics) × 2 (Game Described as Difficult: American government vs. American politics).

Results

Effects of prime on choice of trivia game

As predicted, a logistic regression revealed with marginal significance that money-primed participants chose the game with money in its picture (i.e., the "money game") more often than did control participants (60% vs. 45%), χ^2 ($df = 1$, $N = 127$) = 2.97, $p = .085$. There was also a main effect of Game, reflecting the fact that people preferred American Government (73%) to American Politics (37%), χ^2 ($df = 1$, $N = 127$) = 17.5, $p < .0001$. There was no interaction, χ^2s < 1, $p = .81$.

Misattribution effects on self-knowledge

Liking for challenges. As predicted, when the money game was described as difficult, primed participants reported liking challenges ($M = 6.61$, $SD = 1.61$) more than when the money game was described as easy ($M = 5.56$, $SD = 1.90$), $t(61) = 2.38$, $p = .02$, $d = .60$. Participants in the control group did not show a similar difference, $t < 1$. A 2 (Prime: money vs. control) × 2 (Money Game: easier vs. more difficult) ANOVA revealed the predicted interaction, $F(1, 123) = 4.05$, $p = .05$, $\eta_p^2 = .03$.

Interest in topics of games. When the topic of the money game was American politics, primed reported preferring this topic over American government (M *difference* = 0.66, $SD = 1.21$) more than when the topic of the money game was American government ($M = 0.10$, $SD = 1.81$), but this difference was not significant $t(61) = 1.54$, $p = .15$, $d = .36$. Participants in the control group did not show a similar difference, $t < 1$. The predicted interaction in the 2 (Prime: money vs. control) × 2 (Money Game Topic: politics vs. government) ANOVA was not significant, $F(1, 123) = 1.95$, $p = .16$, $\eta_p^2 = .02$.

The effect of priming on self-knowledge change. Primed participants showed more indication of learning about their preference from their choice than did control participants. That is, primed participants who learned that their chosen game was challenging reported that they preferred challenges more than did primed participants who learned that their chosen game was easy, $t(61) = 2.56$, $p = .01$, $d = .66$ (see means in Table 1). Unprimed participants who learned that their chosen game was challenging did not report that they preferred challenges any more than did unprimed participants who learned that their chosen game was easy, $t < 1$, $d = -.11$ (see Table 1). A 2 (Prime: money vs. control) × 2 (Chosen Game: easier vs. more difficult) ANOVA revealed the predicted interaction, $F(1, 123) = 4.74$, $p = .03$, $\eta_p^2 = .04$.

Table 1
Study 4: Reported challenge liking and choice of tips topic as an effect of money priming and difficultly level of the chosen game.

	Liking challenges		Choosing tips about challenges	
	Money priming	Control	Money priming	Control
Chosen game more difficult	6.69 (1.28)	6.58 (1.70)	0.59 (0.50)	0.47 (0.49)
Chosen game easier	5.56 (2.06)	6.75 (1.43)	0.23 (0.57)	0.46 (0.49)

Note. Standard deviations are in parentheses. Participants rated how much they like challenges on a scale that ranged from *1 = not at all* to *9 = very much*.

Similarly, primed participants showed more interest in the topic of the game they chose. A 2 (Priming: money vs. control) × 2 (Choice: American government vs. American politics) ANOVA on the difference between reported interest in American politics and American government revealed a significant interaction between Priming and Choice, $F(1, 123) = 6.03$, $p = .02$, $\eta_p^2 = .05$. Primed participants who chose American government showed more preference for that game, compared to primed participants who chose American politics, $t(61) = 3.05$, $p < .01$, $d = .88$. By comparison, control participants who chose American government reported no more preference for that topic than did control participants who chose American politics, $t < 1$, $d = .12$. That is, whereas control participants did not prefer one topic to the other, primed participants came to prefer the topic of the game that they were induced to choose by the prime.

Downstream effects: choice of "Tips"

We predicted that primed participants' choice of tips would be guided by their newly-confabulated attributions about their preferences for challenges. Consistent with this prediction, 59% of the primed participants who chose the game that was later revealed as more difficult chose to read tips about how to pursue challenges, in comparison to only 24% of the primed participants who chose the game that was later revealed to be easy, $\chi^2(1, 63) = 5.05$, $p = .005$. There was no such difference among control participants (47% vs. 46%). A 2 (Prime: money vs. control) × 2 (Chosen game: easier vs. more difficult) factorial logistic regression revealed the predicted interaction, χ^2 ($df = 1$, $N = 127$) = 3.98, $p < .05$.

Reported reasons for choice of game

Priming. None of the participants reported that the passage affected their choice of trivia game.

Goal desirability. Unexpectedly, primed participants reported less motivation to earn money ($M = 4.78$, $SD = 2.20$) than control participants ($M = 5.59$, $SD = 2.19$), $t(125) = 2.10$, $p = .04$, $d = .37$. Thus, participants showed no awareness of the effect of priming or of the goal activation.

Reported reasons. On their ratings of the nine potential reasons for their choice of game, the three least important were the cover picture of the game ($M = 2.98$, $SD = 2.33$), the desire to earn money ($M = 2.81$, $SD = 2.45$) and the fact that money appeared on one of the cover pictures ($M = 2.03$, $SD = 1.87$). There was no difference between the ratings of primed and control participants on any of these reasons, $ts < 1$, suggesting that although primed participants were unaware of the effects of the prime. Participants rated their interest in the topic as the most important consideration ($M = 5.52$, $SD = 2.43$), followed by the difficulty of the game ($M = 4.72$, $SD = 2.79$), the desire to relax ($M = 4.58$, $SD = 2.36$), the desire to avoid risks ($M = 4.15$, $SD = 2.28$), and the desire to learn ($M = 3.09$, $SD = 2.26$).

Finally, there was no evidence on the open-ended reason measure that participants were aware of the effect of the money prime on their choice of trivia game. The only participant who mentioned the money on the cover picture as the reason for his choice was in the control group.

Discussion

As predicted, participants who were primed with the goal to earn money were more likely to choose the trivia game that had money on its cover picture, but were unaware of the fact that this goal had affected their choice. Instead, primed participants appear to have misattributed their choice to their dispositional liking of challenges (if the game that they chose was later revealed as challenging) or their dispositional disliking of challenges (if the game they chose was later revealed as easy). These results replicate the findings of Studies 1–3 with an important difference: because participants did not learn about the difficulty level of the game until after they had made their choice, we can say more definitively that primed participants' change in dispositional liking for challenges was a complete confabulation, as opposed to an overattribution to a factor that had some influence on their choice.

Study 4 also established a downstream effect of post-priming misattribution, namely that people's misattributions influenced their subsequent behavior. As predicted, primed participants who learned that their game was challenging were more likely to choose to read tips about how to pursue challenges than were primed participants who learned that their choice of game was easy. This extends the effect of the post-priming misattribution from self-reported preference and traits to actual behavior; even behavior that works against the primed goal that people actually had (i.e., the other option was to get tips about earning money).

General discussion

Four studies established a new type of misattribution effect whereby people acquire faulty self-knowledge: participants in our experiments failed to identify the extent to which a primed goal influenced a choice and attributed that choice to preferences and dispositions unrelated to the goal. As a result, they acquired faulty self-knowledge. In Studies 1 and 2, men primed with the goal to affiliate with women were more likely to choose to interact with a female than a male tutor, but were unaware of the effects of the prime on their choice. Instead, they came to believe that they preferred the topic the female tutor happened to be teaching, and in fact had a dispositional interest in that topic. In Study 3, participants primed with the goal to help others were more likely to choose a word game in which they could help the experimenters (rather than compete with other participants), but were unaware of the effects of the prime on their choice. Instead, they came to believe that they preferred the kind of game (hangman vs. missing letters) that involved helping. In Study 4, participants primed with the goal to earn money were more likely to choose a trivia game associated with money, but again were unaware of the effects of the prime on their choice. Instead, they came to believe that the difficultly level of each game was a reason for their choice.

These findings extend previous research on automatic social behavior by showing its implications for self-knowledge formation. Because social behaviors can be activated automatically without people's awareness, self-knowledge is prone to mistakes. The findings also broaden the scope of self-perception theory by showing that people can misattribute one high-level internal state to another. The present research suggests that even when people's behavior is the result of a high-level mental process, such as the goal to help someone, people are often in "the same position as an outside observer" (Bem, 1972, p. 2) in understanding why they did what they did. Because priming research has shown that many high-level mental processes can be activated automatically and without awareness (Bargh & Ferguson, 2000), we suspect that this post-priming misattribution process is at least as common as the types of misattribution that have been demonstrated previously, namely the misattribution of an external cause of behavior to an internal cause, and the misattribution of the source of physiological arousal.

In such situations it is possible, of course, that people could infer the correct reason for their choice (e.g., infer that the primed goal guided their decision). Participants in the present studies did not make this inference, we suggest, because the primed goals were not accessible or plausible (Nisbett & Wilson, 1977). Because the goal constructs were activated without people's awareness, they were not consciously accessible. The decoy attributes, in contrast, were more accessible and plausible explanations of people's choices.

What did we prime?

One possible question about our studies is whether our priming procedures activated goal pursuit or behaviors related to the primed concept that do not constitute goal pursuit. This distinction has been important in the literature on automaticity, and researchers have taken pains to demonstrate that priming manipulations activated goal pursuit rather than other internal states that led to high-level mental processes (Aarts et al., 2004; Bargh et al., 2001; Custers & Aarts, 2005). Although we used the exact same priming manipulations as many of these studies, this distinction is less important in the present context. For our purposes, the important thing about the priming methods is that they activated high-level mental processes that governed behavior (Bargh & Ferguson, 2000) and had further downstream effects. The purpose of the present work was to show that self-knowledge errors can result from the activation of high-level mental processes via priming (see also Footnote 1).

It is also possible that the primed goals affected behavior because participants attributed the primed concept (e.g., the affiliation goal) to their self-concept (Wheeler, DeMarree & Petty, 2007). That is, the primed goal may have activated a self-concept that fits this goal (e.g., the affiliation-seeking self) and the activation of that self-concept affected subsequent behavior. Regardless of the exact mechanism by which priming influences choices, our studies are concerned with what happens further downstream, namely with people's explanations of their choices that result from priming.

It might also be helpful to clarify the difference between the present research and the process by which people infer that they have caused an effect (Wegner & Sparrow, 2004). When people infer authorship of behavioral outcomes, the components that lead to that attribution are often people's thoughts prior to observing the effect. If the thoughts and the effects match, then authorship is more likely to be attributed to the self. In the present research, the focus was not on people's perceived authorship of an effect; all the behaviors were freely chosen and thus attributed to the self. Rather, the question was how people inferred the reasons that made them choose what they did.

The determinants of choice versus self-attribution

Another possible criticism of our studies is that the awareness probes were insensitive. Although there was little evidence that participants were aware that the primes had increased the accessibility of a goal or influenced their choices, it might be argued that we did not probe carefully enough to reveal such awareness. Though we believe that our probes were adequate, we should emphasize that our hypotheses about self-knowledge acquisition do not depend on people being completely unaware of a goal influencing their behavior. For example, people might have had a fleeting awareness of the goal

but forgot it by the time they made inferences about why they chose what they did.

In principle, people could even be aware that a primed goal influenced them to some degree, but fail to appreciate how much and misattribute their choice to other factors. Indeed, when we directly asked participants how much the primed goal influenced their choice (e.g., how much they were motivated to help in Study 3), some participants reported that it did. Importantly, however, in no study did primed participants rate the primed goal as more influential than did control participants, suggesting that they were not fully aware of how much it influenced their decisions. But even if primed participants had recognized the role of the primed goal to some extent, misattribution could still occur. Research shows that people prefer single over multiple causes (Kelley, 1972; Kruglanski, 1996; Zhang, Fishbach & Kruglanski, 2007), thus the presence of a plausible "decoy" reason might decrease people's belief that a primed goal influenced their behavior. That is, two or more reasons can join together to produce the same choice option, but become competitors in the self-attribution process. One cause of inaccurate self-knowledge, then, may be that people overestimate the influence of some factors at the expense of others.

Another way in which self-knowledge can change after a choice is via post-decisional dissonance reduction (Brehm, 1956), whereby people come to prefer *all* the attributes of the chosen option in order to feel good about their decision. Alternatively, people might increase their liking for all aspects of a chosen alternative because after the choice it is associated with the self ("my choice") and is endowed with the positive value of the self (*associative self-anchoring*, Gawronski, Bodenhausen & Becker, 2007). Though it is possible that these processes occurred in our studies, they did not happen indiscriminately whereby *all* people increased their liking for all aspects of their chosen alternatives. Rather, liking of the attributes of the chosen alternative was often (but not always) stronger among primed participants, in comparison to control participants. In Study 4, primed participants who chose the game that was later revealed as the more challenging game reported more liking of challenges than primed participants who chose the challenging game. Similarly, on average, primed participants also reported a strong preference for the topic of the game that they had chosen. Both these effects did not happen in the control group. Similarly, in Study 2 the choice of the female's tutoring session influenced the self-reported dispositional interest in the topic that she taught more if the participants were primed with the affiliation goal, than if the participants were not primed with that goal. In another three analyses of this difference between control and primed participants in Studies 1 and 2 and Study 3, the pattern of results was the same, but failed to reach statistical significance. When the probabilities of the three non-significant analyses were combined together (Rosenthal, 1978), the combined probability was significant, $z = 2.26$, $p = .02$.

This suggestive evidence does not only decrease the plausibility that, in this research, post-choice dissonance or associative self-anchoring contributed much to the effect of choice on attribution and self-knowledge; it may also suggest that when a mental process influences behavior without people's awareness, they may experience an explanatory vacuum (Oettingen, Grant, Smith, Skinner & Gollwitzer, 2006; Parks-Stamm, Oettingen & Gollwitzer, 2010) that increases their need to explain their own behavior, and by that also increases misattribution and self-knowledge distortion.

Self-knowledge formation

The present research is one instance of a broader framework that identifies hidden causes of behavior and explores the self-(mis) attributions that follow. For example, as noted earlier, similar misattribution effects might occur even if people are initially aware of why they made a decision but later forget those reasons. A frequent example occurs when people step into a room (say, the kitchen) only to realize that they cannot remember the reason that brought them there—inferring, perhaps, that they just have wanted a piece of the cake that happens to be in the refrigerator. Additionally, some goals may be activated with awareness, but their effect on the behavior may seem implausible or self-deprecating. For instance, whereas all researchers can easily detect that they are frustrated when their paper is rejected, some may deny that writing a harsh review about another paper, a few days later, is aimed at anything but promoting good science.

There are several examples in the literature for misattribution of behavior that was caused by self-deprecating reasons. For instance, people who were influenced by the race or gender of job candidates when rating them (e.g., preferred a man over a woman) attributed their choice to reasons other than race or gender (Norton, Vandello & Darley, 2004; Uhlmann & Cohen, 2005). In another example, male participants preferred to sit next to an attractive female in a study only when it was possible to attribute this behavior to reasons other than their romantic interest in her (Bernstein et al., 1983). The present work may suggest that this kind of rationalization is more likely if the self-deprecating reasons affected people behavior without their awareness. Although these lines of work did not focus on people's awareness, it might be the case that the misattribution in those studies was facilitated by participants' low (or lack of) awareness of the actual cause that influenced their behavior.

In closing, we do not mean to imply that people always misattribute choices to the wrong causes. To the extent that the actual cause of a decision is plausible and accessible, people are likely to "discover" the real reason for their choice through a process of self-attribution. People may be especially likely to discover their true goals when they have the opportunity to observe several choices over time. If a college student consistently chooses situations in which she can feel superior to others, for example, she is more likely to attribute those choices to a need to compete.

On the other hand, some pairs of automatically-activated goals and decoy reasons might often co-occur, increasing the likelihood that people internalize an incorrect goal. For instance, people who download songs illegally using file-sharing programs probably do so in order to save money. However, they may construe this habit as a protest against the big record companies' monopolization of the market. Later, these people may support small record companies, even if this support does not save them money, demonstrating internalization of inaccurate self-beliefs. And, once people acquire a faulty theory about why they do what they do, it is particularly difficult for them to observe an actual covariation between their behavior and its potential causes (Nisbett & Ross, 1980)—especially if the actual cause does not fit their image of themselves as good, competent people. Theories about the self may be very much like our theories about the external world—data-based inferences that are often correct but which can go awry.

Acknowledgments

We gratefully acknowledge the support of research grant RO1-MH56075 from the National Institute of Mental Health and a McClelland Center Fellowship awarded to Yoav Bar-Anan.

References

Aarts, H., Custers, R., & Holland, R. W. (2007). The nonconscious cessation of goal pursuit: When goals and negative affect are coactivated. *Journal of Personality and Social Psychology*, *92*, 165−178.

Aarts, H., Gollwitzer, P. M., & Hassin, R. R. (2004). Goal contagion: Perceiving is for pursuing. *Journal of Personality and Social Psychology*, *87*, 23−37.

Bargh, J. A. (2007). Social psychology and the unconscious: The automaticity of higher mental processes. New York: Psychology Press.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, *71*, 230−244.
Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: The automaticity of higher mental processes. *Psychological Bulletin*, *126*, 925−945.
Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trotschel, R. (2001). The automated will: Nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology*, *81*, 1014−1027.
Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside conscious awareness on impression formation. *Journal of Personality and Social Psychology*, *43*, 437−449.
Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, *Vol. 6*. (pp. 1−62)San Diego, CA: Academic Press.
Bernstein, W. M., Stephenson, B. O., Snyder, M. L., & Wicklund, R. A. (1983). Causal ambiguity and heterosexual affiliation. *Journal of Experimental Social Psychology*, *19*, 78−92.
Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *Journal of Abnormal and Social Psychology*, *52*, 384−389.
Cantor, J. R., Zillmann, D., & Bryant, J. (1975). Enhancement of experienced sexual arousal in response to erotic stimuli through misattribution of unrelated residual excitation. *Journal of Personality and Social Psychology*, *32*, 69−75.
Chartrand, T. L., & Bargh, J. A. (1996). Automatic activation of impression formation and memorization goals: Nonconscious goal priming reproduces effects of explicit task instructions. *Journal of Personality and Social Psychology*, *71*, 464−478.
Chartrand, T., Dalton, A., & Fitzsimons, G. (2007). Nonconscious relationship reactance: When significant others prime opposing goals. *Journal of Experimental Social Psychology*, *43*, 719−726.
Custers, R., & Aarts, H. (2005). Positive affect as implicit motivator: On the nonconscious operation of behavioral goals. *Journal of Personality and Social Psychology*, *89*, 129−142.
Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology*, *74*(4), 865−877.
Dutton, D. G., & Aron, A. P. (1974). Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology*, *30*, 510−517.
Eitam, B., Hassin, R. R., & Schul, Y. (2008). Nonconscious goal pursuit in novel environments: The case of implicit learning. *Psychological Science*, *19*, 261−267.
Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of Personality and Social Psychology*, *41*, 232−242.
Fazio, R., Zanna, M., & Cooper, J. (1977). Dissonance and self-perception: An integrative view of each theory's proper domain of application. *Journal of Experimental Social Psychology*, *13*, 464−479.
Ferguson, M. J., Hassin, R., & Bargh, J. A. (2008). Implicit motivation: Past, present, and future. In J. Shah, & W. Gardner (Eds.), *Handbook of motivational science* (pp. 150−166). New York: Guilford.
Festinger, L., & Carlsmith, J. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203−210.
Fishbach, A., & Labroo, A. A. (2007). Be better or be merry: How mood affects self-control. *Journal of Personality and Social Psychology*, *93*, 158−173.
Fitzsimons, G. M., & Bargh, J. A. (2003). Thinking of you: Nonconscious goals associated with relationships partners. *Journal of Personality and Social Psychology*, *84*, 148−164.
Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, *4*, 195−202.
Fries, A., & Frey, D. (1980). Misattribution of arousal and the effects of self-threatening information. *Journal of Experimental Social Psychology*, *16*, 405−416.
Gawronski, B., Bodenhausen, G. V., & Becker, A. P. (2007). I like it because I like myself: Associative self-anchoring and post-decisional change of implicit attitudes. *Journal of Experimental Social Psychology*, *43*, 221−232.
Gazzaniga, M. (1985). The social brain: Discovering the networks of the mind. New York: Basic Books.
Gazzaniga, M. S., & LeDoux, J. E. (1978). The integrated mind. New York: Plenum Press.
Gorassini, D. R., & Olson, J. M. (1995). Does self-perception change explain the foot-in-the-door effect? *Journal of Personality and Social Psychology*, *69*, 91−105.
Hassin, R. R., Bargh, J. A., & Zimerman, S. (2009). Automatic and flexible: The case of non-conscious goal pursuit. *Social Cognition*, *27*, 20−36.
Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.). (2005). *The new unconscious*. New York: Oxford University Press.
Higgins, E. T., Rholes, W. S., & Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, *13*, 141−154.
Holland, R. W., Hendriks, M., & Aarts, H. (2005). Smells like clean spirit: Nonconscious effects of scent on cognition and behavior. *Psychological Science*, *16*, 689−693.
Kelley, H. H. (1972). Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 151−174). Morristown, NJ: General Learning Press.
Kiesler, C. A., Nisbett, R. E., & Zanna, M. P. (1969). On inferring one's beliefs from one's behavior. *Journal of Personality and Social Psychology*, *11*, 321−327.
Kruglanski, A. W. (1996). Goals as knowledge structures. In P. M. Gollwitzer, & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 599−619). New York: Guilford Press.
Loersch, C., Aarts, H., Payne, B. K., & Jefferis, V. E. (2008). The influence of social groups on goal contagion. *Journal of Experimental Social Psychology*, *44*, 1555−1558.
Olson, J. M. (1990). Self-inference processes in emotion. In J. M. Olson, & M. P. Zanna (Eds.), *Self-inference processes: The Ontario Symposium*, *Vol. 6*. (pp. 17−41)Hillsdale, NJ: Erlbaum.
Nisbett, R. E., & Ross, L. D. (1980). Human inference: Strategies and shortcomings of human inference. Englewood Cliffs, NJ: Prentice Hall.
Nisbett, R. E., & Valins, S. (1972). Perceiving the causes of one's own behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 63−78). New York: General Learning Press.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231−259.
Norton, M. I., Vandello, J. A., & Darley, J. M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology*, *87*, 817−831.
Oettingen, G., Grant, H., Smith, P. K., Skinner, M., & Gollwitzer, P. M. (2006). Nonconscious goal pursuit: Acting in an explanatory vacuum. *Journal of Experimental Social Psychology*, *42*, 668−675.
Parks-Stamm, E. J., Oettingen, G., & Gollwitzer, P. M. (2010). Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving. *Journal of Experimental Social Psychology*, *46*, 531−542.
Ramachandran, V. S. (1996). The evolutionary biology of self deception, laughter, dreaming and depression: Some clues from anosognosia. *Medical Hypotheses*, *47*, 347−362.
Rosenthal, R. (1978). Combining the results of independent studies. *Psychological Bulletin*, *85*, 185−193.
Schachter, S., & Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, *69*, 379−399.
Shah, J. Y., & Kruglanski, A. W. (2003). When opportunity knocks: Bottom-up priming of goals by means and its effects on self-regulation. *Journal of Personality and Social Psychology*, *84*, 1109−1122.
Shariff, A. F., & Norenzayan, A. (2007). God is watching you: Priming God concepts increases prosocial behavior in an anonymous economic game. *Psychological Science*, *18*, 803−809.
Sheeran, P., Webb, T. L., & Gollwitzer, P. M. (2005). The interplay between goal intentions and implementation intentions. *Personality and Social Psychology Bulletin*, *31*, 87−98.
Uhlmann, E., & Cohen, G. L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science*, *16*, 474−480.
Wegner, D. M., & Sparrow, B. (2004). Authorship processing. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1201−1209), 3rd Edition. Cambridge, MA: MIT Press.
Wheeler, S. C., DeMarree, K. G., & Petty, R. E. (2007). Understanding the role of the self in prime-to-behavior effects: The active-self account. *Personality and Social Psychology Review*, *11*, 234−261.
White, G., Fishbein, S., & Rutstein, J. (1981). Passionate love and the misattribution of arousal. *Journal of Personality and Social Psychology*, *41*(1), 56−62.
Wilson, T. D. (2002). Strangers to ourselves: Discovering the adaptive unconscious. Cambridge, MA: Harvard University Press.
Zanna, M., & Cooper, J. (1974). Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology*, *29*, 703−709.
Zhang, Y., Fishbach, A., & Kruglanski, A. W. (2007). The dilution model: How additional goals undermine the perceived instrumentality of a shared path. *Journal of Personality and Social Psychology*, *92*, 389−401.