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