European Journal of Social Psychology, *Eur. J. Soc. Psychol.* **44**, 255–266 (2014)
Published online 21 February 2014 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/ejsp.2005

Research article

Confabulating reasons for behaving bad: The psychological consequences of unconsciously activated behaviour that violates one's standards

MARIEKE A. ADRIAANSE*, JONAS WEIJERS, DENISE T. D. DE RIDDER, JESSIE DE WITT HUBERTS AND CATHARINE EVERS
Department of Clinical and Health Psychology, Utrecht University, Utrecht, The Netherlands

Abstract

Numerous studies have been conducted to demonstrate that behaviours are frequently activated unconsciously. The present studies investigate the downstream psychological consequences of such unconscious behaviour activation, building on work on the explanatory vacuum and post-priming misattribution. It was hypothesized that unconsciously activated behaviours trigger a negative affective response if the behaviour violates a personal standard and that this negative affect subsequently motivates people to confabulate a reason for the behaviour. Results provided evidence for this mediated moderation model. Study 1 showed that participants who were primed to act less prosocially indeed reported increased levels of negative affect and, as a result, were inclined to confabulate a reason for their behaviour. Study 2 replicated these findings in the domain of eating and provided evidence for the moderating role of personal standards as well as the entire mediated moderation model. These findings have relevant theoretical implications as they add to the modest number of studies that demonstrate that the effect of unconscious priming may extend well beyond performing the primed behaviour itself to influence subsequent affect and attribution processes.

Until the 1990s, it was commonly assumed that people make conscious decisions on which goals to strive for and which actions to engage in. Over the past 20 years, however, nothing short of a revolution has taken place in psychological research, with numerous studies showing that traits, concepts, and goals can also be activated outside of conscious awareness to influence behaviour. For example, it has been found that priming people with the trait rudeness can cause them to interrupt others more quickly and more frequently (Bargh, Chen, & Burrows, 1996; Study 1) and priming people with the African-American stereotype can increase hostility (Bargh et al., Study 3). Unconscious primes may even steer people in a direction that is opposite to a chronic goal or standard. For example, research on the goal conflict model of eating shows that when cues from the environment (temptations) trigger the unconscious goal of eating enjoyment in chronic dieters, the competing conscious goal of weight control becomes temporarily inhibited (Stroebe, Mensink, Aarts, Schut, & Kruglanski, 2008; Stroebe, Van Koningsbruggen, Papies, & Aarts, 2013). As a result of this hedonic priming, these chronic dieters violate their weight control goal by indulging in palatable foods without having an explanation for doing so.

Relative to the vast amount of studies that have demonstrated that a large number of our daily behaviours are initiated unconsciously (Bargh & Chartrand, 1999) and that unconsciously activated behaviours may sometimes be unwanted (e.g. overeating while trying to control one's weight), research investigating the downstream psychological consequences, or 'after-effects' of unconsciously activated behaviours, is however short in supply (c.f. Chartrand, Cheng, Dalton, & Tesser, 2010). In the present study, we aim to build on recent research on the affective consequences of acting in an 'explanatory vacuum' (Oettingen, Grant, Smith, Skinner, & Gollwitzer, 2006; Parks-Stamm, Oettingen, & Gollwitzer, 2010) and research on 'post-priming misattribution' (Bar-Anan, Wilson, & Hassin, 2010) to shed more light on these downstream consequences. More specifically, on the basis of the aforementioned research, we aim to test a proposed sequence of events that may occur when people have been primed to act in a way that conflicts with their personal standards. First of all, on the basis of the work by Oettingen et al. and Parks-Stamm et al., we argue that the realization that one has acted to violate one's standards without having an explanation for this (i.e. acting in an 'explanatory vacuum'; Oettingen et al; Parks-Stamm et al.) will trigger a negative affective response. However, inspired by recent insights from Parks-Stamm et al. and Bar-Anan et al., we hypothesize that this negative affect is not the endpoint, but rather the beginning of a new behavioural sequence motivating people to *confabulate*—to make a false claim without the intent to deceive and without knowing that this claim is ill-grounded (Hirnstein, 2009)—a plausible explanation for their unwanted behaviour. Notably, we expect that such confabulations are less likely to be made when

*Correspondence to: Marieke A. Adriaanse, Department of Clinical and Health Psychology, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands.
E-mail: M.A.Adriaanse@uu.nl

Received 14 February 2013, Accepted 21 January 2014

the behaviour does not violate a personal standard (Oettingen et al.; Parks-Stamm et al.).

The specific aims of the present study are as follows: (i) to replicate the finding that priming participants to violate a certain personal standard increases negative affect (Oettingen et al., 2006; Parks-Stamm et al., 2010); (ii) to replicate the finding that, in case of unconsciously activated behaviour that violates a personal standard, people confabulate a reason for this behaviour (Bar-Anan et al., 2010; Parks-Stamm et al.); (iii) to test for the first time the proposition that negative affect *fuels* this confabulation process, or, in other words, to provide the first evidence that negative affect *mediates* the relation between performing a primed behaviour and confabulation; and finally, (iiii) to bring the aforementioned steps together in one model. Specifically, we hypothesize that a mediated moderation model (Muller, Judd, & Yzerbyt, 2005) applies: Both the direct and indirect (via negative affect) effects of performing an unconsciously activated behaviour on confabulation are expected to be moderated by personal standards (Figure 1), such that negative affect and confabulation only occur in case the primed behaviour violates a personal standard.

Evidence for the first step of the proposed sequence of events—that unawareness of the cause of one's behaviour leads to negative affect when this behaviour violates personal standards—was first provided by Oettingen et al. (2006). Whereas other research (Chartrand & Bargh, 2002) had demonstrated that the affective consequences of conscious versus unconscious goal pursuits are similar, with success in unconscious goal striving leading to a positive and failure leading to negative 'mystery mood', Oettingen et al. hypothesized that effects may be different when successful goal striving involves behaviour that violates a norm. Specifically, on the basis of research showing that accidental harmdoing leads to more guilt than intentional harmdoing (McGraw, 1987), Oettingen and colleagues hypothesized that behaviour that violates a prevailing norm and that is the result of a nonconsciously, but not consciously, activated goal will trigger negative affect. In other words, the authors argued that when people act in an 'explanatory vacuum', that is, when behaviour cannot be explained by a conscious goal or normative explanation and thus *demands* an explanation, negative affect will arise.

To test this hypothesis, participants in the Oettingen et al. (2006) study were asked to participate in a collaborative task with a partner. Participants either consciously or unconsciously adopted the goal to be accommodating (a goal that conformed to the social norm to cooperate) or combative (a goal that violated the social norm to cooperate) or, in the control condition, no goal was adopted. Results were in line with the predictions: After acting in line with their goal, only participants who were unconsciously primed with the combative (norm-violating) goal showed an increase in negative affect. These results thus suggest that unconsciously, but not consciously, provoking people to perform a certain behaviour triggers negative affect, but only if this behaviour is norm violating and thus requires an explanation.

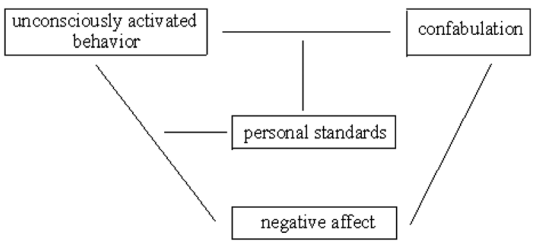

Figure 1. The proposed sequence of events including (a) a direct effect of performing unconsciously activated behaviour on confabulation that is moderated by personal standards and (b) an indirect effect of enacting an unconsciously activated behaviour on confabulation via negative affect that is moderated by personal standards

The present research aims to replicate the finding that priming participants to violate a certain norm, or standard, increases negative affect. However, we also assume that this negative affective reaction is not the endpoint. Rather, on the basis of the work by Parks-Stamm et al. (2010) and Bar-Anan et al. (2010), we expect that negative affect functions as a *motivational force* driving people to confabulate a reason for their behaviour. Specifically, Parks-Stamm et al. (2010) followed up on the research by Oettingen et al. (2006), arguing—in line with research on cognitive dissonance (e.g. Stone & Cooper, 2001)—that people who have been unconsciously provoked to act in a way that violates their personal standards 'should be motivated to find an alternative explanation for their behaviour, thereby reducing the negative affect associated with an explanatory vacuum' (p. 532). This assumption was based on the notion that negative affect—resulting from an inexplicable discrepancy between one's behaviour and one's standards—is an aversive state that people are motivated to reduce. Indeed, Parks-Stamm et al. demonstrated that participants who were primed to act in a way that violates a salient norm felt significantly less negative when this behaviour was in line with a previously provided conscious goal. This finding implies that people are likely to 'use' an available plausible explanation in order to reduce the negative affect that is associated with the lack of such an explanation.

Other evidence for the proposition that confabulation is likely to occur following unconscious behaviour activation has been provided by Bar-Anan et al. (2010). In one of their studies, it was shown that priming participants with the goal to look for female companionship caused male participants to choose a course that was given by a female instructor more often than a course given by a male instructor, regardless of the course's actual topic. Importantly, however, participants believed that the course's topic was the most important reason for their choice. In another study, the authors demonstrated confabulation even more convincingly by showing that participants misattributed a choice to a cue that was actually provided *after* actually making the choice. In this study, participants who were primed with a goal to earn money were more likely to later prefer a game with pictures of American presidents as they appear on dollar bills over another game that depicted normal pictures of the same American presidents. It was only after indicating their preference for either one of these games that participants received information about the games' difficulty. In this way, the participants could not have used this information to choose their game. Still, participants who received information that their game was difficult reported that they liked difficult games more than participants who later learned that their game of choice was easy, clearly demonstrating that participants were confabulating.

Although Bar-Anan et al. (2010) thus provided convincing evidence for the notion that confabulation is a likely consequence of unconscious behaviour activation, their studies notably did not concern behaviours that violate a norm. However, one explanation for why confabulation still occurred may be that Bar-Anan et al. focused on what may be called *provoked* confabulation (Berlyne, 1972), which occurs when being explicitly asked to explain the behaviour. As the likelihood of causal attribution increases in case of an explicit request to explain one's behaviour (Wilson, Dunn, Kraft, & Lisle, 1989), probing for an explanation that is in reality not accessible is likely to promote confabulation regardless of whether a behaviour is norm violating or not. However, when people are not directly probed to provide an explanation for their behaviour (a more ecologically valid situation), as is the case in the present study, we argue that people are only expected to confabulate when the unconsciously activated behaviour violates a certain standard and thus demands an explanation.

The proposed sequence of events, which will be tested in two studies, is thus mainly inspired by research on the explanatory vacuum (Oettingen et al., 2006; Parks-Stamm et al., 2010) and research on post-priming misattribution (Bar-Anan et al., 2010). However, the present studies also complement this work in the sense that they for the first time directly test whether negative effect *mediates* the effect of unconsciously activated behaviour on confabulation. In the studies by Bar-Anan et al., affect was not assessed. In the Parks-Stamm et al. studies, affect was assessed, but mediation could not be tested. That is, the goal to which participants could attribute their behaviour was provided *before* participants were primed to act in a norm-violating manner. Results showed that when a previous goal could explain the behaviour, negative affect did not increase, suggesting, according to Parks-Stamm and colleagues, that participants had attributed their behaviour to this goal. An advantage of the design employed in this study is that it allowed the authors to confirm their hypothesis that confabulation occurs reflexively. However, a disadvantage of this method is that mediation could not be tested. In the present study, we changed the timing for offering the information that participants can use for confabulations (after performing the unwanted unconsciously activated behaviour and after the assessment of negative affect), so that mediation can be formally tested. Lastly, and most importantly, in the present study, we aim to bring together all of the aforementioned steps in one model and hypothesize that all steps combined are best described in terms of a *mediated moderation model* (Muller et al., 2005). That is, we expect that both the direct and indirect (via negative affect) effects of performing an unconsciously activated behaviour on confabulation are moderated by personal standards (Figure 1).

With the present studies, we aim to shed more light on the downstream consequences of unconscious behaviour activation. Although the applications and mechanisms of unconscious behaviour activation have received, and continue to receive, a great deal of attention in psychological research, relatively few studies have investigated the psychological consequences of unconsciously activated behaviour, such as consequences for mood (but see Bongers, Dijksterhuis, & Spears, 2009; Chartrand & Bargh, 2002; Oettingen et al., 2006; Parks-Stamm et al., 2010) or misattribution (but see Bar-Anan et al., 2010; Parks-Stamm et al., 2010).

Yet, it should also be acknowledged that the proposed sequence of events outlined in Figure 1 has some strong ties to classic research, in particular to research on cognitive dissonance (Elliot & Devine, 1994; Festinger, 1957; Stone & Cooper, 2001). Quite similar to the first step outlined in the confabulation model, research on cognitive dissonance has shown that people experience discomfort (dissonance) when an inconsistency exists between two cognitions, most notably one's behaviour and one's attitude (Festinger, 1957), or, according to the more recent Self-standards Model (Stone & Cooper, 2001), between a person's behaviour and a person's self-concept. However, whereas in the present study, it is hypothesized that people subsequently become inclined to attribute their unconsciously activated behaviours to other plausible causes, empirical work on cognitive dissonance has focused on attitude change as the means to reduce dissonance.[1] In fact, in a typical dissonance experiment, researchers actually make an explicit effort to ensure that no other plausible explanations are available as people only resort to changing their attitudes when other plausible explanations (and thus possibilities for confabulation) are lacking (Festinger & Carlsmith, 1959; Gosling, Denizeau, & Oberle, 2006; Kiesler & Pallak, 1976; Zanna & Cooper, 1974).

Although cognitive dissonance and the proposed confabulation process are thus similar in terms of the mediating role of negative affect and the general process of justification, these two processes also depart from each other in proposing two separate 'routes of justification' (i.e. confabulation or classical dissonance reduction through attitude change) to account for behaviours that are inconsistent with one's self-standards. In addition to proposing different routes of justification, it should also be noted that unlike in priming experiments, in dissonance studies, participants are never completely unaware of the cause of their behaviour (usually a request by the experimenter). In other words, rather than having insufficient justification for one's behaviour as is the case in cognitive dissonance studies, in the present studies, people have, similar to other priming studies, no justification for their behaviour.

The Present Research

Two studies were conducted to test our hypotheses. Study 1 was designed to test the first three assumptions: (i) that being unconsciously primed to act less prosocially triggers negative affect; (ii) that being unconsciously primed to act less prosocially triggers a tendency to confabulate; and (iii) that negative affect mediates the relation between the unconsciously activated behaviour and confabulation. Study 2 aimed to replicate the findings from Study 1 and to extend them by investigating the confabulation model in another domain (eating) and by investigating the moderating role of personal

[1]However, note that the original Dissonance Theory by Festinger, as well as it successors, such as the New Look Model (Cooper & Fazio,1984) and the Self-standards Model (Stone & Cooper, 2001), did theoretically allow for alternative ways of reducing dissonance that do not involve attitude change.

standards. Moreover, Study 2 was designed to also explicitly address our fourth aim, that is, to test the entire proposed mediated moderation model.[2]

STUDY 1

In Study 1, participants received a neutral or antisocial priming manipulation and were then asked to participate in an 'extra' experiment for a fellow student without getting additional reimbursement. Participants in the antisocial prime condition were expected to quit sooner on this extra task compared with participants in the neutral condition. In line with our hypotheses, quitting sooner without having a plausible explanation for this was expected to lead to increased levels of negative affect and a subsequent need to confabulate a reason for this behaviour.

Previous research has shown that people who have a high desirability for control have a stronger 'illusion of control', that is, they have a stronger tendency to perceive personal control over outcomes that are in fact not under their control (Burger & Cooper, 1979). To make sure that any results could not be attributed to differences in desirability for control across the two conditions, the Desirability of Control Scale (Burger, 1992) was administered at baseline. For similar reasons, several items assessing participants' helpfulness and liking of computer games were included in a baseline questionnaire. Note that these items were purposely 'hidden' among many filler items and framed in general terms to avoid jeopardizing the credibility of the cover story regarding the 'extra task'.

Method

Participants

Sixty-one university students were invited to participate in a study on gaming in exchange for €5 or course credit. Because of computer errors, two participants could not complete the entire study so that the final sample included 59 students (21 men, 37 women, and one person who did not indicate his or her sex) with a mean age of 21.25 years ($SD = 3.68$).

[2]Before moving on to the first study, it is important to provide a clear description of what is meant by 'unconscious behaviour activation' in the present paper. For unconsciously activated behaviours, there may be several aspects of which people are unaware. Specifically, Chartrand (2005) identified three types or stages of awareness: awareness of the environmental features that trigger an automatic process, the automatic process itself, and the outcome of that automatic process. When individuals are unaware of one or more of these stages, the process can be considered unconscious. In the proposed research, the situation of interest is one in which the crucial aspect of unawareness is the automatic process itself. The case of interest for the present paper is one in which an individual may or may not be aware of the environmental features triggering the automatic process (note that this also implies that priming manipulations may be subliminal or supraliminal) and is aware of the behaviour they are performing, but unaware of the process by which this behaviour has come about (i.e. unaware of the influence of the prime). Also, as we aim to provide a broad working theory for the downstream consequences of nonconscious behaviour, in the proposed model, we do not discriminate between priming goals to activate unconscious goal pursuit or priming other concepts that active behaviour that is not necessarily goal directed (c.f. Bar-Anan et al., 2010).

Design

The experiment had a factorial design with two conditions (neutral prime vs antisocial prime)[3].

Procedure

Participants were told that they would be participating in a study on gaming. After filling out several questionnaires including a baseline measure (T0) of negative affect, participants were randomly assigned to play either one of two computer games for 10 minutes, which was selected to serve as a neutral prime or an antisocial prime (see priming manipulation). After completing this task and before moving on to the final part of the experiment, participants were asked to participate in an 'extra' task. It was explained that by participating in this extra task, they would really help out a fellow student but that they would not receive any reimbursement for this extra task. Participants all commenced with this task, but they were told that whenever they felt they had completed enough trials, they were free to quit. After quitting the extra task, participants continued with the focal experiment, which at that point involved filling out a second measure of negative affect (T1). Participants were then informed that the study was now finished but that it would be appreciated if they could answer some questions about the new lab facilities that they had just used. It was hypothesized that participants in the antisocial condition would be more negative about the lab facilities, or in other words, that they would use these questions as a plausible confabulation justifying their less prosocial behaviour (i.e. completing fewer trials of the 'extra' task). Finally, participants were debriefed, thanked, and paid.

Materials

Baseline questionnaire. At baseline (T0), participants were asked to indicate their sex, age, and current level of education. Then, a so-called lifestyle questionnaire was administered, which included 30 items about several domains (e.g. health). The items were rated on 7-point scales ranging from (1) *totally not applicable to me* to (7) *totally applicable to me*. All of the items (e.g. 'I eat sufficient fruit and vegetables') were filler items, except for three items that pertained to playing (computer) games (e.g. 'I enjoy playing computer games') and three items that assessed helpfulness ('I think that everyone should look after themselves'). These six items were included in the randomization check. After the lifestyle questionnaire, the Desirability of Control Scale (Burger, 1992) was administered. The scale includes 20 statements (e.g. 'I enjoy having control over my own destiny', $\alpha = .82$) rated on 7-point Likert scales ranging from (1) *totally not applicable to me* to (7) *totally applicable to me*. Lastly, participants filled out the 20 items of the Positive and Negative Effect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). For the present

[3]In view of recent concerns raised by Simmons, Nelson, and Simonsohn (2011), we deemed it relevant to disclose that participants who came to the lab were randomized to one of four conditions. That is, in addition to the two conditions reported in the present paper (to which 61 students were assigned), which were designed to test the hypotheses outlined in the introduction, two additional conditions designed to test a different research question were run by students at the same time for educational purposes.

study, we were interested in the negative affect subscale of this questionnaire (PANAS_NA) in order to allow for controlling for potential baseline differences in negative affect. The PANAS_NA includes 10 items to measure negative affect (T0: $\alpha = .81$) on 5-point scales ranging from (1) *not at all* to (5) *very much*.

Priming manipulation. The priming manipulation was based on meta-analytical evidence (e.g. Anderson, 2004; Anderson & Bushman, 2001) showing that aggressive games lead to less prosocial, or helping, behaviour (also Ferguson, 2007). In the present study, Tetris was the neutral game and Lamers the antisocial game. Tetris involves moving and rotating blocks that fall down a vertical shaft in such a manner that they create horizontal line of blocks without gaps. The goal of the Lamers game is to kill as many Lemmings as possible using a machine gun, a pistol, a bomb, and a mine, before getting to the gate at the other side. A pilot study among 78 students was conducted to test whether participants who had just played Lamers were less helpful in an unrelated next task (i.e. participating in an extra study for fellow students without reimbursement) compared with participants who had just played Tetris. Results showed that indeed, compared with Tetris, Lamers was found to lead to less helping behaviour on the subsequent task, $F(1, 76) = 6.19$, $p = .02$.

Helping behaviour. After finishing playing the computer game, participants commenced with an ostensibly unrelated extra experiment that was ran by fellow students and for which they would not receive any reimbursement. Participants were told that this involved a continuous task and that the students required as many trials as possible but that every extra completed trial would be very helpful. It was explained that in each trial, a box with a hole on one side would be presented very briefly and that they were supposed to indicate on which side the box had a hole by pressing the corresponding arrow key. The experimenter waited outside the cubicle and informed participants that they should open the door of the cubicle if they felt that they had helped enough. As soon as the door was opened, the experimenter entered the cubicle, closed the experiment, and thanked the participant on behalf of the students. The dependent variable was the number of completed trials. As this variable was skewed, it was transformed using a square root transformation before being entered in the analyses. However, to facilitate interpretation, means and standard deviations are reported for the nontransformed variable.

Follow-up questionnaire. After finishing the extra task, the PANAS was administered for the second time. Again, we were interested in the PANAS_NA (T1: $\alpha = .81$).

Confabulation. After filling out the PANAS, participants were told that the experiment was now finished but that we would appreciate some feedback on the new lab facilities they had just used. This cover story was deemed highly credible as the studies were indeed conducted in newly built labs. To this end, participants received the Lab Evaluation Questionnaire with several questions about the lab that could serve as *post hoc* rationalizations (i.e. confabulation) for why they quit early on the 'extra' task. Specifically, participants were asked to respond to nine questions on 5-point scales, ranging from (1) *not at all* to (5) *very much*, that assessed participants' evaluations of several aspects of the lab, such as the comfortableness of the chair and the temperature in the labs (e.g. 'the chair was comfortable'; $\alpha = .78$). This task thus served to assess whether the participants in the experimental condition would confabulate that their decreased tendency to help was a result of the lab facilities. Positively framed items were recoded so that a higher score indicated a more negative evaluation of the lab, and thus a stronger tendency to confabulate.

Awareness Check

Approximately 83% of the participants were available[4] for an awareness check. They were asked the following: (i) what they thought was the purpose of this study and (ii) whether they noticed anything about the game they had played (Tetris or Lamers). The large majority (88%) indicated no awareness whatsoever. Six participants opted that the extra task may have been part of the experiment, but of those participants, only three participants indicated that they thought that the game and the extra task might have been related. These participants however did not correctly identify how (i.e. type and/or direction of the effect) the game and the extra task were related. As rerunning the main analyses without excluding these six or three participants yielded similar results, all participants were retained in the analyses.

Results

For means, standard deviations, and intercorrelations of the key variables under study, please see Table 1.

Randomization Check

Separate analyses of variance (ANOVAs) with age, baseline negative affect, the three questions about gaming, the three questions about helpfulness and desirability of control as dependent variables and condition as the independent variable were conducted to test whether randomization was successful. A chi-square analysis was conducted to test whether sex was equally distributed across the two conditions. None of the results were significant, all p's $> .21$, indicating successful randomization.

Manipulation Check

To examine whether helping behaviour was successfully manipulated, an ANOVA with condition as the independent variable and the number of trials completed in the extra task as the dependent variable was conducted. The ANOVA yielded a significant effect of condition, $F(1, 57) = 5.69$, $p = .02$, $\eta_p^2 = .09$, indicating that participants in the antisocial condition completed significantly fewer trials ($M = 88.90$, $SD = 72.95$) than participants in the neutral condition ($M = 133.53$, $SD = 89.40$).

Negative Affect

A repeated-measures ANOVA with time as a within-subject variable, condition as a between-subjects variable, and negative

[4]Owing to time constraints, not all participants were available for the awareness check.

Table 1. Study 1: means, standard deviations, and correlations

	1	2	3	4	5	6	7
Sex (1)	—						
Age (2)	−0.11	—					
Desirability of control (3)	−0.04	0.09	—				
Negative affect T0 (4)	−0.17	−0.10	0.01	—			
Number of trials (5)[a]	0.05	0.16	−0.20	−0.08	—		
Negative affect T1 (6)	0.03	−0.19	0.01	0.58*	−0.19	—	
Confabulation (7)	−0.02	−0.14	−0.24	0.17	−0.17	0.29*	—
M	21[b]	21.25	4.69	1.38	111.59	1.42	2.18
SD	—	3.68	0.65	0.42	84.09	0.43	0.58

[a]Correlations are reported for the transformed variable, but *M* and *SD* are reported for the nontransformed variable.
[b]Percentage of male participants.
*$p < .05$.

affect as the dependent variable yielded a nonsignificant effect of time, $F < 1$, and a significant Time × Condition interaction, $F(1, 57) = 4.91$, $p = .03$, $\eta_p^2 = .08$. Follow-up repeated-measures ANOVAs in the two conditions separately showed that negative affect marginally significantly increased from baseline ($M = 1.41$, $SD = .47$) to follow-up ($M = 1.57$, $SD = .49$) in the antisocial condition, $F(1, 28) = 3.92$, $p = .058$, $\eta_p^2 = .12$, but showed a nonsignificant decrease from baseline ($M = 1.34$, $SD = 0.35$) to follow-up ($M = 1.28$, $SD = 0.33$) in the neutral condition, $p = .30$.

Confabulation

An ANOVA with condition as a between-subjects variable and lab evaluation as the dependent variable was used to determine whether participants in the antisocial condition would be inclined to attribute their lower helping behaviour to attributes of the lab. Indeed, a significant effect of condition was found, $F(1, 57) = 5.42$, $p = .02$, $\eta_p^2 = .09$, with participants in the antisocial condition providing a more negative evaluation about the lab ($M = 2.35$, $SD = 0.58$) compared with participants in the neutral condition ($M = 2.01$, $SD = 0.54$).

Mediation

To test whether the effect of condition on negative affect (T1) mediated the effect on lab evaluation, three subsequent regression analyses were conducted according to the guidelines of Baron and Kenny (1986). In line with the results presented earlier, the first two regression analyses showed that the relationship between condition and lab feedback was significant, $\beta = .30$, $p = .02$, as was the relationship between condition and negative affect (T1), $\beta = .33$, $p = .01$. However, when negative affect was included in the third regression analysis, the relationship between condition and lab feedback decreased in strength and was no longer significant, $\beta = .23$, $p = .10$, indicating that negative affect partially mediates the relation between condition and lab evaluation (i.e. confabulation).

Alternative Explanation

The effects on negative affect and confabulation could arguably have been caused by the behaviour itself (i.e. the number of trials completed). To test more strictly whether negative affect and confabulation are indeed the result of performing an unwanted behaviour *without having an explanation for doing so*, rather than performing an unwanted behaviour in itself (i.e. irrespective of whether or not one has introspection into the reasons for this behaviour), additional analyses were performed: It was tested whether the effects on negative affect and confabulation are still present when controlling for the behaviour (i.e. the number of trials). If the effects on negative affect and confabulation are entirely driven by the unwanted behaviour on its own, the effects on negative affect and confabulation should disappear when adding the number of trials completed as a covariate.

When controlling for the number of trials completed, the results for negative affect and confabulation remained the same. The only exception was that negative affect now even significantly reduced from T0 to T1, $F(1, 28) = 4.72$, $p = .04$, $\eta_p^2 = .14$, in the neutral condition (and still increased in the antisocial condition, $F(1, 28) = 6.57$, $p = .02$, $\eta_p^2 = .20$), providing even more support for our assumptions. These findings indicate that the previously reported effects on negative affect and confabulation were due to the fact that participants in the antisocial condition quitted sooner without having an explanation for this, rather than merely quitting sooner itself.

Discussion

Results from Study 1 replicated previous findings that acting in an explanatory vacuum triggers negative affect and a tendency to confabulate (Bar-Anan et al., 2010; Parks-Stamm et al., 2010). Specifically, Study 1 showed that people who were unconsciously primed to act less prosocially experienced an increase in negative affect after quitting sooner on the extra task and subsequently confabulated by devaluating the lab environment. Importantly, these findings also extended earlier work by showing that the elevated levels of negative affect mediated the effect on confabulation. Including the number of trials completed as a covariate did not change the pattern of results. It can therefore be concluded that negative affect and the corresponding need to confabulate arise as a result of quitting sooner without having an explanation for this (i.e. acting in an explanatory vacuum), and not as a result of quitting sooner *per se*. The present study thus supports the hypothesis that helping less as a result of unconscious priming triggers negative affect and that this negative affect functions as a motivational force fuelling the need to confabulate a reason for this behaviour.

STUDY 2

The second study was designed to replicate the results of Study 1 in another domain and to extend findings from Study 1 by investigating the hypothesized moderating role of personal standards. In doing so, Study 2 endeavoured to test the whole model as outlined in Figure 1 using a formal mediated moderation analysis. In this model, confabulation is specified as the dependent variable, and it is assumed that personal standards moderate the relation between the independent variable (performing an unconsciously activated behaviour) and confabulation as well as the relation between the mediator (negative affect) and confabulation. That is, an increase in negative affect and confabulation are only expected to occur when the primed behaviour violates personal standards. We tested the entire mediated moderation model according to the steps described by Muller et al. (2005).

Method

Participants

Sixty-three female university students participated in exchange for €4 or course credit. After excluding one participant who did not follow the instructions (i.e. did not consume any chocolates during the taste test; Procedure section), the final sample consisted of 62 participants with a mean age of 21.45 years ($SD = 3.41$) and a mean body mass index of 22.36 ($SD = 3.62$).

Design

The study had a 2 (dieting standard: high vs low) × 2 (condition: neutral prime vs hedonic prime) between-subjects design.[5]

Procedure

Participants were told that they were participating in three independent studies: a study on concentration and word recognition that included a lexical decision task (LDT), a taste test, and a 'text comprehension task'. Different fonts and layouts were used in the questionnaires, and different names were printed on the covers of the questionnaires to enforce our cover story that they were part of different studies. In reality, the LDT served to prime participants in the experimental condition with food enjoyment, the taste test served to measure food intake, and the text comprehension task served to assess confabulation.

As part of the alleged first study, participants filled out a bogus questionnaire on factors affecting alertness (to enforce the cover story), a measure of affect, and a so-called lifestyle questionnaire in which participants' dieting standards were unobtrusively assessed. Participants then completed an LDT aimed at priming participants with food enjoyment words (experimental condition) or neutral words (control condition). Participants then moved on to the alleged second study, which involved participating in a 'taste test' in which we unobtrusively measured the effect of the priming manipulation on the grammes of chocolates consumed. Finally, they started with the third study, in which affect was measured for the second time and in which they performed a 'text comprehension task'. In this task, participants had to highlight the key sentences of a scientific article that explained that depletion of concentration increases cravings for sugar.

Participants were then told that they were finished with all three studies but that it would be appreciated if they could give some feedback on the studies they had just participated in. Specifically, we asked them to judge how hard each task was and how much concentration it required. This task thus served to measure whether the participants in the experimental condition, after reading the article on the relation between concentration depletion and sugar cravings, would confabulate that their consumption of chocolates was a result of the level of concentration needed for the LDT that they did earlier.

Materials

Questionnaires. At baseline, participants reported some demographics and answered six questions about factors that could influence responsiveness (e.g. 'I use soft-drugs') to enforce the cover story of the first study. To measure negative affect, the PANAS was administered. Again, we were interested in the PANAS_NA (T0: $\alpha = .84$). In addition, baseline hunger and thirst were assessed on scales ranging from (1) *not at all* to (5) *very much*. Lastly, a so-called lifestyle questionnaire was administered, which included 10 filler items (e.g. 'I smoke') and one item to inconspicuously determine dieting standards ('I try to restrict my snack intake') rated on scales ranging from (1) *not at all* to (7) *very much*.

Lexical decision task. In order to unconsciously prime half of the participants with the concept of food enjoyment, participants did an LDT, which was highly similar to one that previously primed food enjoyment successfully (Fishbach, Friedman, & Kruglanski, 2003; Stroebe et al., 2008). Each trial contained a string of letters that was briefly presented (23 ms) and that was preceded and followed by a pre-mask and post-mask (a row of x's; 500 ms). The letter string that followed could either be an existing word or a nonexisting word, and it was up to the participants to determine whether the letter string was a word or not by pressing a 'yes' or 'no' button as quickly as possible. The LDT consisted of 100 trials that were presented in random order. In the experimental condition, there were 14 critical trials with hedonic target words (e.g. 'delicious'). In the control condition, these critical target words were replaced by neutral words (e.g. 'table'). The other trials were the same in both conditions: the target letter strings were either neutral words (36) or nonwords (50).

Taste test. The second task was the taste test. Participants were given 200 g of chocolate to taste. The experimenter explicitly mentioned that they could eat as much as they liked and could take their time as it would take at least 15 minutes before they could start with the next task. Participants were asked to respond to 15 questions about their appreciation of the chocolates (e.g. regarding flavour, texture, or colour) using 7-point scales ranging from (1) *not at all* to (7) *very much* as

[5]For exploratory reasons, the experiment included one additional condition in which participants were primed with health-related words. This prime proved unsuccessful (participants in the health and the control condition consumed an equal amount of chocolates, $p = .91$). For reasons of legibility and conciseness, this condition is not included in the present analyses.

well as one open-ended question that involved describing the package they would design for this product. Unbeknownst to the participants, before and after the taste test, the contents of the bowl of chocolates were weighed to determine how much the participants had eaten (cf., Evers, De Ridder, & Adriaanse, 2009). The dependent variable was the amount of chocolates consumed in grammes. As this variable was skewed, it was transformed using a square root transformation before being entered in the analyses. However, to facilitate interpretation, means and standard deviations are reported for the nontransformed variable.

Text comprehension task. The alleged third experiment was introduced as a study on emotions and text comprehension. Participants started with filling out the PANAS for the second time (PANAS_NA T1: $\alpha = .82$). In reality, this measure of negative affect served to assess whether participants in the experimental condition felt more negative affect after violating their dieting standards compared with participants in the control condition. Participants were then asked to read an article concerning the relationship between glucose and concentration. As a cover story, participants were told that the purpose of this task was to assess how Dutch students deal with reading English texts because universities increasingly assign English texts to students. Participants were told to carefully read the text and to highlight what they thought were the five most important sentences of the article.

We used a slightly adapted excerpt (1.5 pages) of an article about a paper by Gailliot et al. (2007), which appeared in the Association for Psychological Science magazine *The Observer* (January 2009). Adaptations to the summary from *The Observer* were minor and mainly involved ensuring that the article was fully understandable for our student sample. The article explains that exerting self-control, for example by doing a boring, but cognitively demanding task, depletes glucose levels so that people become inclined to restore this by consuming sugar.

Confabulation. To assess whether participants used the content from the article to confabulate a reason for their behaviour, they were asked to fill out one last questionnaire that was said to function as a general feedback form but that essentially measured how demanding participants found the LDT. The evaluation of the LDT involved seven items, which could be answered on 7-point scales ranging from 1 (*totally disagree*) to 7 (*totally agree*) and which asked about depleting qualities of the tasks (e.g. 'the task was too long to really stay concentrated' and 'the task was demanding'; $\alpha = .81$).

Debriefing

Before debriefing, participants were asked the following: (i) whether they had noticed something about the target words in the LDT; (ii) what they thought was the purpose of the LDT; and (iii) what they thought was the purpose of the experiment in general. Without probing, no comments were given about the priming manipulation in particular. After specifically asking whether the participants noticed anything particular about the LDT, several participants from both conditions opted that some of the tasks may have been related, but none of them discerned the true purpose of the experiment. Still, to make sure that possible awareness of the priming manipulation did not influence our findings, all analyses that are reported later were also conducted without these particular participants. As the results remained similar, all participants were retained in the analyses.

Results

For means, standard deviations, and intercorrelations of the key variables under study, please see Table 2.

Randomization Check

To check whether randomization was successful, separate ANOVAs were conducted with condition as the independent variable and with age, body mass index, baseline hunger, baseline thirst, baseline negative affect, and dieting standard as dependent variables. No significant effects were found (all p's $> .24$).

Manipulation Check

An ANOVA with condition as a between-subjects variable and the grammes of chocolates consumed as the dependent variable was used to determine whether the priming manipulation was successful. Condition had a significant effect on the amount

Table 2. Study 2: means, standard deviations, and correlations

	1	2	3	4	5	6	7	8	9
Age (1)	—								
BMI (2)	0.18	—							
Hunger (3)	−0.18	−0.09	—						
Thirst (4)	−0.08	0.22	0.28*	—					
Dieting Standards (5)	−0.01	−0.05	0.10	−0.08	—				
Negative Affect T0 (6)	−0.19	−0.15	0.18	0.29*	0.22	—			
Chocolates consumed (grammes) (7)[a]	−0.23	−0.08	0.36*	0.12	0.04	0.06	—		
Negative Affect T1 (8)	−0.11	−0.24	0.30*	0.35*	0.18	0.76*	0.10	—	
Confabulation (9)	−0.18	−0.41*	0.30*	0.05	−0.11	0.27	0.13	0.40*	—
M	21.45	22.36	2.16	2.76	4.44	1.40	31.40	1.33	4.44
SD	3.41	3.62	1.09	1.25	1.44	0.47	27.69	0.40	0.98

[a]Correlations are reported for the transformed variable, but *M* and *SD* are reported for the nontransformed variable.
*$p < .05$.

of chocolates consumed, $F(1, 60) = 4.33$, $p = .04$, $\eta_p^2 = .07$, with participants in the hedonic condition eating approximately 68% more ($M = 39.73$, $SD = 34.07$) than those in the neutral condition ($M = 23.59$, $SD = 17.09$).

To ensure that dieting standards did not influence the effectiveness of the priming manipulation, a hierarchical linear regression was conducted with the amount of chocolates consumed as the dependent variable and condition, dieting standards, and an interaction term of both as predictors. Variables were mean centred before being entered into the regression analysis. Results showed that condition was a significant predictor, $\beta = .26$, $p = .046$, whereas dieting standards and the interaction term were not (p's $> .79$), indicating that the effectiveness of the priming manipulation was not affected by dieting standards.

Testing the Model

In order to test whether the entire model could be validated, a mediated moderation analysis was carried out, following the procedure described by Muller et al. (2005). This procedure involves performing three consecutive regression analyses. In all regression analyses, independent variables were mean centred before computing the interaction terms.

The first regression analysis was designed to determine the effect of the independent variable (condition), the moderator (dieting standards), and an interaction term of both on the dependent variable (confabulation). The model proved to be significant, $F(3, 58) = 2.90$, $p = .04$. Of the independent variables, the main effects of condition ($p = .32$) and dieting standards ($p = .45$) were not significant. The interaction term of Condition × Dieting Standards was significant, $\beta = .32$, $p = .01$, suggesting an overall moderating effect of the relationship between condition and confabulation by dieting standards. According to Muller et al. (2005), the lack of a main effect of condition ($p = .32$) and the presence of a significant Condition × Dieting Standard interaction is indicative of mediated moderation.

Before conducting the other two regression analyses to further substantiate this suggestion, the observed interaction effect was further explored. To examine this interaction, simple slopes were computed for participants with low versus high dieting standards (+1 *SD* vs −1 *SD* of the mean dieting standard score: Aiken & West, 1991). Simple slopes analyses indicated that when participants had relatively low dieting standards, condition had no significant effect on confabulation ($\beta = -.20$, $p = .26$), but when participants had relatively high dieting standards, condition had a significant effect on confabulation ($\beta = .45$, $p = .01$). This suggests that only when dieting standards are high does acting in an explanatory vacuum cause confabulation.

The second regression analysis was designed to determine the effect of the independent variable (condition), the moderator (dieting standards), and an interaction term of both on the mediator (negative affect). The results show that the model as a whole was significant, $F(3, 58) = 4.09$, $p = .01$. Again, the effects of condition ($p = .15$) and dieting standards ($p = .11$) were not significant, but the effect of the interaction term Condition × Dieting Standards on negative affect was significant, $\beta = .33$, $p < .01$. In order to facilitate interpretation, simple slopes were computed for participants with low versus high dieting standards. Simple slopes analyses indicated that when participants had relatively low dieting standards, condition had no significant effect on negative affect ($\beta = -.16$, $p = .34$), but when participants had relatively high dieting standards, condition had a significant effect on negative affect ($\beta = .52$, $p < .01$). This suggests that only when dieting standards are high does acting in an explanatory vacuum cause negative affect.

Lastly, a regression analysis was conducted with dieting standards, condition, an interaction term of Condition × Dieting Standards, negative affect, and an interaction term of Dieting Standards × Negative Affect as independent variables and confabulation as the dependent variable. The results show that the model as a whole was significant, $F(5, 56) = 3.74$, $p < .01$, with a significant main effect of negative affect on confabulation, $\beta = .36$, $p < .01$, and a marginally significant effect of the Condition × Dieting Standards interaction term on confabulation, $\beta = .23$, $p = .08$. The other effects were insignificant, all p's $> .25$.

Results showed that the moderating effect by dieting standards was reduced from $\beta = .32$ to a marginally significant effect of $\beta = .23$ after controlling for negative affect and the interaction of Negative Affect and Dieting Standards. According to Muller et al. (2005), this is indicative of mediated moderation and can be interpreted as negative affect partially mediating the moderation effect of Dieting Standards on the relation between condition and confabulation.

Alternative Explanation

The effects on negative affect and confabulation, reported in the previous section, could arguably have been caused by violating one's dieting standards in itself (i.e. the amount of chocolates consumed), and not by acting in an explanatory vacuum. Similarly to Study 1, we therefore replicated the previous analyses while controlling for the amount of chocolates consumed. All results remained the same. This suggests that the effects were not caused by violation of dieting standards *per se*, but rather by violating one's dieting standards as a result of being unconsciously primed with hedonic words.

Discussion

Study 2 replicated the findings from Study 1 regarding negative affect and the subsequent need to confabulate. However, this study also extends findings from Study 1 as similar results were obtained in another behavioural domain and, in line with Oettingen et al. (2006), personal standards were now included as a moderator. Taken together, Study 2 provides evidence for the process by which acting in an explanatory vacuum triggers confabulation (i.e. negative affect) and the condition under which confabulation is most likely to occur (i.e. when violating personal standards). Most importantly, however, Study 2 provided evidence for the proposed mediated moderation model.

GENERAL DISCUSSION

Two studies were presented that together provide convincing evidence for the proposed mediated moderation model

describing the downstream consequences of unconscious behaviour activation (Figure 1). We replicated the finding that in the aftermath of unconsciously activated behaviour, people experience negative affect (Oettingen et al., 2006; Parks-Stamm et al., 2010) and become inclined to confabulate a reason for their behaviour (Bar-Anan et al., 2010; Parks-Stamm et al., 2010). As expected, effects on negative affect and confabulation were most likely when the primed behaviour violated personal standards (Oettingen et al., 2006; Parks-Stamm et al., 2010). Notably, results from both studies also extended previous work as they provided evidence for the assumption that negative affect functions as a mediator, or a motivational force, driving the confabulation process. Moreover, results from Study 2 showed for the first time that these downstream consequences are best described by a mediated moderation model with personal standards moderating the relation between acting in an explanatory vacuum and confabulation, as well as the relation between negative affect (the mediator) and confabulation. Importantly, negative affect and confabulation appeared not to be the result of the personal standard violation itself, but rather a consequence of the fact that participants violated a personal standard without having an explanation for doing so. Taken together, the present findings have relevant theoretical implications as they add to the modest number of studies that demonstrate that the effect of unconsciously activated behaviours may extend well beyond performing the primed behaviour itself to influence cognitions, beliefs, and evaluations, such as beliefs about ourselves (e.g. beliefs about our tendency to consume sugar when performing a cognitively demanding task).

In addition to cognitive dissonance (Festinger & Carlsmith, 1959; Zanna & Cooper, 1974; see introduction), the present findings have ties with several other classic research traditions, such as research on motivated reasoning (Kunda, 1990), which posits that 'people motivated to arrive at a particular conclusion attempt to be rational and to construct a justification of their desired conclusion that would persuade a dispassionate observer' (p. 484), and with research indicating that people rely on observable cues to deduce the reasons for their behaviour (Bem, 1972). However, the present findings also go beyond these studies as they experimentally investigated the use of such justifications in the context where it should arguably be most prevalent, but where attribution processes have, surprisingly, remained relatively under-investigated (c.f. Bar-Anan, 2010): the context of unconscious behaviour activation where people by definition have no access to the actual cause of their behaviour. The present findings also strongly relate to the proposition by Wegner (Wegner, 2002; Wegner & Wheatley, 1999; Wilson, 2002) that people have an 'illusion of conscious will', meaning that in most cases where behaviour is activated outside of conscious awareness or by forces other than our own will, people still assume that they have somehow 'willed' this behaviour (for an illustrative study demonstrating this phenomenon, see Wegner & Wheatly, 1999). Rather than accepting the status of acting in an explanatory vacuum, people thus appear to be inclined to fill this gap by looking for explanations for why they 'decided' to act in a way that violates their standards.

Finally, our findings indicating that confabulation is a likely consequence of unconscious behaviour activation are in line with classic work by Nisbett and Wilson (1977), which showed that people generally have little ability to accurately report on their cognitive processes and consequently form *post hoc* causal theories to explain their behaviour. Our studies, however, also go beyond their findings in investigating the circumstances supporting as well as the process leading to these confabulations. Moreover, similar to Bar-Anan (2010), Nisbett and Wilson (1977) focused on *provoked* confabulation (Kopelman, 1980), which is elicited specifically in response to questions that probe [participants'] memory (Hirnstein, 2009, p. 92). The likelihood of causal attribution increases in case of an explicit request to explain one's behaviour (Wilson et al., 1989). In the present studies, we specifically tried to avoid making an explicit request to explain the unconsciously activated behaviour in order to allow for testing whether confabulation processes may also occur relatively spontaneously. In this sense, a more ecologically valid situation is created where people are not directly probed to provide an explanation for their behaviour.

Although the present studies address a more spontaneous form of confabulation in comparison with provoked confabulation where participants are explicitly requested to provide an explanation for their behaviour, still it has to be noted that in the present studies, the opportunity for confabulation was provided explicitly to the participants. That is, without providing people with the presently used confabulations, it is unlikely that participants would have spontaneously explained their behaviour in terms of the laboratory environment (Study 1) or glucose deficits (Study 2). A related issue concerns the degree to which people are consciously aware that they are confabulating. Whereas the present studies cannot answer this question, findings by Parks-Stamm et al. (2010) suggest that confabulation occurs reflexively, without conscious awareness: In these studies, need for cognition and the possibility for conscious reflection did not moderate any of the effects, suggesting that people have an automatic tendency to reduce the negative affect associated with acting in an explanatory vacuum.

Two limitations of the present studies need to be addressed. First, in both studies, negative affect was assessed using self-report measures. Because self-reports are inherently limited, future research might benefit from the addition of implicit measures of negative affect (e.g. Quirin, Kazen, & Kuhl, 2009). By using implicit measures, affect could be measured directly without the possible biases that may occur when using self-reports. Second, the use of a college student sample can be considered a limitation as it remains unclear whether the tendency to confabulate a reason for unconsciously activated behaviour is equally strong for other samples.

Despite these limitations, the present findings yield convincing evidence for the proposed mediated moderation model and add to recent insights (most notably by Bar-Anan et al, 2010; Oettingen et al., 2006; Parks-Stamm et al., 2010) into what happens in the aftermath of unconsciously activated behaviour. These findings could potentially have important implications as they suggest that priming has relevant downstream psychological consequences that extend beyond the behaviour that was triggered. Moreover, whereas (more severe) confabulation is frequently studied in clinical samples, for example among split-brain patients or patients with Korsakoff syndrome (e.g. Borsutzky, Fujiwara, Brand, & Markowitsch, 2008), less is known about the processes underlying and circumstances

leading to confabulation among healthy individuals, despite a general consensus that confabulation also occurs in healthy people (for an overview, see Hirnstein, 2009). The present findings thus add to the literature on 'non-clinical confabulation' by providing evidence on the circumstances under which confabulation can occur and by exploring in more depth the process leading towards confabulation as a result of unconscious behaviour activation among healthy individuals. Future research should be conducted to investigate how stable confabulated reasons are over time and how they might influence subsequent behaviour.

Future research is also required to investigate under which circumstances people are more inclined to confabulate a reason for their behaviour (thus keeping their standards intact) or resort to classic cognitive dissonance reduction by changing their standards. It seems likely that changing attitudes or personal standards is more probable in case of more ambiguous or weak attitudes. Moreover, research on attribution has demonstrated that in general people are most inclined to attribute their behaviour to reasons that are accessible, plausible (Bar-Anan-2010; Nisbett & Wilson, 1977; see also research on the representativeness and availability heuristic, Tversky & Kahneman, 1973) and self-promoting (Kunda, 1990), implying that confabulation might be a more likely consequence following unconsciously activated norm-violating behaviour, the more accessible, plausible or self-promoting explanations are available. These suggestions, however, remain to be investigated in future research.

CONCLUSION

Notwithstanding the limitations of the present studies and the questions that are still to be investigated, the current research adds to the literature on the after-effects of unconsciously activated behaviours. Building on recent work on the explanatory vacuum (Oettingen et al., 2006; Parks-Stamm et al., 2010) and post-priming misattribution (Bar-Anan et al., 2010), we showed that there are some relevant effects that occur only after performing a primed behaviour. Specifically, in line with previous work, we showed that acting in an explanatory vacuum gives rise to negative affect and confabulation. Most importantly, however, we showed that this sequence of events occurring as a result of acting in an explanatory vacuum can be explained by a mediated moderation model: Both the direct and indirect effects of performing an unconsciously activated behaviour on confabulation are moderated by personal standards; negative affect and confabulation only arise when a primed behaviour violates a personal standard. Future research is required to investigate how these confabulations affect subsequent behaviour and when confabulation takes precedence over other routes of justification, most notably, classic cognitive dissonance reduction by means of changing one's attitudes or standards.

ACKNOWLEDGEMENT

The research in this paper was supported by a grant (VENI-451-11-030) from the Netherlands Organization for Scientific Research, awarded to the first author.

REFERENCES

Aiken, L. S., & West, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Anderson, C. A. (2004). An update on the effects of playing violent video games. *Journal of Adolescence*, *27*, 113–122.

Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behaviour, aggressive cognition, aggressive affect, physiological arousal, and prosocial behaviour: A meta-analytic review of the scientific literature. *Psychological Science*, *12*, 353–359.

Bar-Anan, Y., Wilson, T. D., & Hassin, R. R. (2010). Inaccurate self-knowledge formation as a result of automatic behaviour. *Journal of Experimental Social Psychology*, *46*, 884–894.

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

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behaviour: Direct effects of trait construct and stereotype priming on action. *Journal of Personality and Social Psychology*, *71*, 230–244.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.

Bem, D. J. (1972). Self-perception theory. *Advances in Experimental Social Psychology*, *6*, 1–57.

Berlyne, N. (1972). Confabulation. *British Journal of Psychiatry*, 120, 31–39. In Hirnstein, W. (2009). *Confabulation: views from neuroscience, psychiatry, psychology and philosophy*. Oxford: Oxford University Press.

Bongers, K. C. A., Dijksterhuis, A., & Spears, R. (2009). Self-esteem regulation after success and failure to attain unconsciously activated goals. *Journal of Experimental Social Psychology*, *45*, 468–477.

Borsutzky, S., Fujiwara, E., Brand, M., & Markowitsch, H. J. (2008). Confabulations in alcoholic Korsakoff patients. *Neuropsychologia*, *46*, 3133–3143.

Burger, J. M. (1992). *Desire for control: Personality, social, and clinical perspectives*. New York: Plenum Press.

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion*, *3*, 381–393.

Chartrand, T. L. (2005). The role of conscious awareness in consumer behaviour. *Journal of Consumer Psychology*, *15*, 203–210.

Chartrand, T., & Bargh, J. A. (2002). Nonconscious motivations: Their activation, operation, and consequences. In A. Tesser, D. A. Stapel, & J. W. Wood (Eds.), *Self and motivation: Emerging psychological perspectives* (pp. 13–41). Washington, DC: APA.

Chartrand, T. L., Cheng, C. M., Dalton, A. N., & Tesser, A. (2010). Nonconscious goal pursuit: Isolated incidents or adaptive self-regulatory tool? *Social Cognition*, *28*, 569–588.

Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, *17*, (pp. 229–245). Orlando, FL: Academic Press.

Elliot, A. J., & Devine, P. G. (1994). On the motivational nature of cognitive dissonance: Dissonance as psychological discomfort. *Journal of Personality and Social Psychology*, *67*, 382–394.

Evers, C., De Ridder, D. T. D., & Adriaanse, M. A. (2009). Assessing yourself as an emotional eater: Mission impossible? *Health Psychology*, *28*, 717–725.

Ferguson, C. J. (2007). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior*, *12*, 470–482.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston: Row, Peterson.

Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203–210.

Fishbach, A., Friedman, R. S., & Kruglanski, A. W. (2003). Leading us not into temptation: Momentary allurements elicit overriding goal activation. *Journal of Personality and Social Psychology*, *84*, 296–309.

Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., & Tice, D. M. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, *92*, 325–336.

Gosling, P., Denizeau, M., & Oberle, D. (2006). Denial of responsibility: A new mode of dissonance reduction. *Journal of Personality and Social Psychology*, *90*, 722–733.

Hirnstein, W. (2009). *Confabulation: Views from neuroscience, psychiatry, psychology and philosophy*. Oxford: Oxford University Press.

Kiesler, C. A., & Pallak, M. S. (1976). Arousal properties of dissonance manipulations. *Psychological Bulletin*, *83*, 1014–1025.

Kopelman, M. D. (1980). Two types of confabulation. Neurology, neurosurgery, and psychiatry. In W. Hirnstein (2009) (Ed.), *Confabulation: Views from neuroscience, psychiatry, Psychology and philosophy*. Oxford: Oxford University Press.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*, 480–498.

McGraw, K. M. (1987). Guilt following transgression: An attribution of responsibility approach. *Journal of Personality and Social Psychology*, *53*, 247–256.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

Oettingen, G., Grant, H., Smith, P. K., Skinner, M., & Gollwitzer, P. M. (2006). Unconscious 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 unconscious goal striving. *Journal of Experimental Social Psychology*, *46*, 531−542.

Quirin, M., Kazen, M., & Kuhl, J. (2009). When nonsense sounds happy or helpless: The Implicit Positive and Negative Affect Test (IPANAT). *Journal of Personality and Social Psychology*, *97*, 500–516.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allow presenting anything as significant. *Psychological Science*, *11*, 1359–1366.

Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, *37*, 228–243.

Stroebe, W., Mensink, W., Aarts, H., Schut, H., & Kruglanski, A. W. (2008). Why dieters fail: Testing the goal conflict model of eating. *Journal of Experimental Social Psychology*, *24*, 26–36.

Stroebe, W., Van Koningsbruggen, G. M., Papies, E. K., & Aarts, H. (2013). Why most dieters fail but some succeed: A goal conflict model of eating behavior. *Psychological Review*, *120*(1), 110–138.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–232.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

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

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

Wilson, T. D. (2002). *Strangers to ourselves: Discovering the adaptive unconscious*. Cambridge, MA: Harvard University Press.

Wilson, T. D., Dunn, D. S., Kraft, D., & Lisle, D. J. (1989). Introspection, attitude change, and attitude-behaviour consistency: The disruptive effects of explaining why we feel the way we do. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. *22*, pp. 287–343). Orlando, FL: Academic Press.

Zanna, M. P., & 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.