RESEARCH ARTICLE

Explaining unexplainable food choices

Marieke A. Adriaanse*, Floor M. Kroese*, Jonas Weijers*, Peter M. Gollwitzer†,‡ & Gabriele Oettingen†,§

* Utrecht University, Utrecht, The Netherlands
† New York University, New York City, New York, USA
‡ University of Konstanz, Konstanz, Germany
§ University of Hamburg, Hamburg, Germany

Correspondence
Marieke A. Adriaanse, Department of Clinical & Health Psychology, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands.
E-mail: m.a.adriaanse@uu.nl

Received: 24 August 2015
Accepted: 21 October 2016

http://dx.doi.org/10.1002/ejsp.2273

Keywords: nonconscious, explanatory vacuum, confabulation, eating, attribution

Abstract

In recent years, psychologists have started to investigate the downstream consequences of nonconsciously activated behaviour (acting in an 'explanatory vacuum'). Results have shown that when such behaviour is norm-violating, people experience a need to confabulate reasons for this behaviour. The present paper aims to add more convincing evidence for this assumption. Study 1 addresses this question by replicating Study 2 of Adriaanse, Weijers, De Ridder, De Witt Huberts, and Evers (2014) while adding a condition in which people are post hoc provided with an explanation for their behaviour. Study 2 addresses this question by explicitly demanding an explanation for a nonconsciously steered choice. Both studies were conducted in the context of eating behaviour. Results of both studies were indicative of confabulation as a downstream consequence of nonconsciously steered eating behaviour (Study 1) or food choice (Study 2). Future research should address the potential of confabulated reasons spilling over to next occasions.

Explaining Unexplainable Food Choices

People generally believe that their behaviour is a consequence of their conscious will (Wegner & Wheatley, 1999). For example, people may refer to factors like 'motivation' or 'persistence' as important determinants of success in pursuing their goals, such as eating less unhealthily. However, this belief that the realization of our goals is determined by our consciously willed actions stands in stark contrast with evidence showing that a large part of our daily behaviour (e.g., our health-related behaviours, Sheeran, Gollwitzer, & Bargh, 2013) is activated outside of conscious awareness. For example, studies have shown that when habitual drinkers are primed with the goal of socializing, they automatically become more inclined to drink alcohol (Sheeran et al., 2005), and dieters who are primed with palatable foods automatically inhibit the goal to control their weight (Stroebe, Mensink, Aarts, Schut, & Kruglanski, 2008). Considering that people believe that they have conscious control over their behaviour, psychologists have recently started to wonder how people deal with nonconsciously activated behaviour when this behaviour is norm-violating[1] (e.g., indulging in palatable foods while on a diet). Initial findings on this topic suggest that nonconsciously activated norm-violating behaviour triggers a tendency to 'confabulate' (make up) plausible reasons for this behaviour (e.g., Adriaanse, Weijers, De Ridder, De Witt Huberts, & Evers, 2014; Parks-Stamm, Oettingen, & Gollwitzer, 2010). The present paper aims to provide more convincing evidence for this recent proposition.

The notion that nonconsciously activated norm-violating behaviour triggers a tendency to confabulate was first proposed by Oettingen and colleagues (Oettingen, Grant, Smith, Skinner, & Gollwitzer, 2006; Parks-Stamm et al., 2010). Whereas most research focused on demonstrating the similarities between conscious and nonconscious goal pursuit, Oettingen and colleagues (Oettingen et al., 2006; Parks-Stamm et al., 2010) theorized that although the outcome of both types of goal pursuit may indeed be similar, the experience of goal pursuit may differ depending on whether the resulting actions are in line with, or violate, prevailing norms. Where in case of conscious goal pursuit, the norm-violating behaviour can be explained by referring to the conscious intentions, in case of nonconscious goal pursuit such an explanation is—by definition—lacking. Building on the notion that experiencing an inconsistency between one's behaviour and normative standards feels aversive and motivates attempts to restore consistency (e.g., Stone & Cooper,

[1]Please note that in the present paper, we refer to norm-violation in the wider sense, including behaviours that violate all kinds of consciously held standards, including goals, attitudes or (social) norms.

2001), it was argued that in case of nonconsciously activated norm-violating behaviour, actors should become motivated to confabulate an alternative explanation for their behaviour in order to reduce the negative affect accompanying this 'explanatory vacuum' (Oettingen et al., 2006; Parks-Stamm et al., 2010). The term confabulation was used to describe this post hoc attribution process, as this signifies that people adopt an erroneous reason for their behaviour 'without the intent to deceive and without knowing that this claim is ill-grounded' (Hirstein, 2009). This aligns with the classic work by Nisbett and Wilson (1977) demonstrating that people have limited introspective awareness and as a result form post hoc causal theories to explain their behaviour without actually being aware of this interpretative process (Parks-Stamm et al., 2010).

In a first series of studies testing their theory, Oettingen et al. (2006) indeed found that acting on nonconscious goals that trigger norm-violating behaviour increases negative affect; acting on respective conscious goals was not associated with an increase in negative affect. In a second series of studies, it was found that participants primed to perform a norm-violating behaviour did not demonstrate this increase in negative affect when a previous conscious goal, which was in fact unrelated to the norm-violating behaviour, could explain this behaviour (Parks-Stamm et al., 2010). This finding was considered to be the first evidence to suggest that people acting in an explanatory vacuum are motivated to confabulate a reason for their behaviour. Inspired by the findings of Parks-Stamm and colleagues, Adriaanse et al. (2014) proposed and found support for a model describing these downstream consequences of acting in an explanatory vacuum. According to this model, people will experience negative affect when nonconsciously activated behaviours violate salient personal norms or standards (i.e., personal standards act as a moderator). This negative affective reaction, in turn, operates as a motivational force (i.e., a mediator) directing an individual to confabulate a reason for the behaviour. The two studies they conducted confirmed their hypotheses. For example, it was found that students who were primed to quit early in an experiment of a fellow student, thereby violating the norm to be helpful, experienced more negative affect and subsequently confabulated that the lab facilities were uncomfortable.

The studies by Adriaanse et al. (2014) provided a more direct test of confabulation as compared with the study by Parks-Stamm et al. (2010), where the absence of negative affect was interpreted as an indication of confabulation. However, in the studies by Adriaanse and colleagues, confabulation was still only inferred from ratings on questionnaires. For example, confabulation was inferred from the fact that participants, who were primed to act less prosocially in the lab, subsequently rated the lab facilities as less comfortable or from the fact that dieters primed to indulge in palatable foods, subsequently rated a lexical decision task (LDT) that preceded their food intake as more cognitively demanding (after they were unobtrusively exposed to this excuse). While it makes sense to assume that these ratings were reflective of an attempt to post hoc rationalize their norm-violating behaviour, an alternative hypothesis that cannot be ruled out is that participants were simply more negative in their ratings of the lab facilities/the LDT, because they felt more negative at the moment they provided these ratings.

The present paper aims to extend recent insights showing that nonconsciously activated norm-violating behaviour triggers attempts to confabulate a reason for this behaviour (e.g., Adriaanse et al., 2014; Parks-Stamm et al., 2010). Specifically, the present research aims to provide more convincing evidence for the notion that previous findings on confabulation can indeed be interpreted as the individual's attempts to explain nonconsciously provoked behaviour rather than reflecting a mere mood congruency bias. To this end, two studies were conducted. Study 1 addresses this question by replicating Study 2 of Adriaanse et al. (2014) while adding a condition in which people are post hoc provided with an explanation for their behaviour. If previous results are indeed indicative of an attempt to explain the behaviour then participants who have just received an explanation in this additional experimental condition should not score high on the confabulation questionnaire. Study 2 addresses this question by explicitly demanding an explanation. Both studies are conducted in the context of eating behaviour as this is a behaviour known to be influenced by automatic processes and mindless decisions (e.g., Wansink & Sobal, 2007).

In sum, the present studies provide more stringent tests of a novel theory describing what happens in the aftermath of nonconsciously activated behaviour. Relative to the number of studies demonstrating the prevalence and working mechanisms of nonconscious goal pursuit, studies on the downstream psychological consequences of nonconscious goal pursuit are scarce, warranting a more thorough investigation of these after-effects (cf. Chartrand, Cheng, Dalton, & Tesser, 2010). In addition, the present studies extend previous work on the limits of introspection (Nisbett & Wilson, 1977) and cognitive dissonance theory (Festinger, 1957) to the domain of nonconscious goal pursuit.

Study 1: Explaining Unexplainable Indulgence

Study 1 was designed as a replication and extension (the extension involved the inclusion of the 'hedonic prime and tell' condition, see below) of a previous study by Adriaanse et al. (2014, Study 2). In Study 1, the effects of a 'neutral prime' condition versus a 'hedonic prime' condition or a 'hedonic prime and tell' condition on chocolate consumption and the subsequent experience of negative affect and confabulation were investigated. To manipulate norm-violation, the study included people with either high or low dieting standards, assuming that eating a lot of chocolates is norm-violating for people with high dieting standards. The procedure of

the study involved a priming manipulation to prime people with neutral or hedonic words (hedonic prime and hedonic prime and tell condition) and a subsequent taste test to unobtrusively measure chocolate intake. The crucial difference between the hedonic prime condition and the hedonic prime and tell condition was that participants in the latter condition were, after completing the taste test, provided with an explanation for their behaviour by telling them that the words that had been presented during the priming task could have resulted in an increased chocolate consumption.

After the taste test, in a supposedly unrelated subsequent task, participants were provided with a scientific article explaining that concentrating on cognitively demanding tasks increases cravings for sugar. Confabulation was assessed by measuring to what extent participants, after reading this text, retrospectively reported that the priming task (i.e., the task proceeding the taste test) was cognitively exhausting. Including the hedonic prime and tell condition allows to rule out the possibility that norm-violation per se, regardless of whether the reason for the norm-violating behaviour is known, would yield higher scores on the confabulation questionnaire. Ruling this out would increase our confidence that higher scores on the confabulation questionnaire for participants in the hedonic prime condition are indeed indicative of confabulation.

Method

Participants. One hundred and twenty-nine female university students participated in exchange for 6 euro or course credit. Due to computer problems, one participant could not finish the experiment. After excluding three participants who did not adhere to the instructions (they did not consume any chocolates in the taste test), the final sample consisted of 125 participants with a mean age of 20.85 years ($SD = 2.77$) and a mean body mass index (BMI) of 22.12 kg/m^2 ($SD = 3.70$).

Design. The study had a 2 (dieting standard: high vs. low) × 3 (condition: neutral prime vs. hedonic prime vs. hedonic prime and tell) between subjects design. The main dependent variable was confabulation. This was operationalized as the degree to which participants indicated that they were cognitively exhausted (as a potential explanation for their indulgent behaviour, see Materials).

Procedure. The general procedure was the same as Adriaanse et al. (2014, Study 2), with the addition of a hedonic prime and tell condition. Participants thought that they participated in three unrelated studies: a 'word recognition and concentration study', a 'taste test' and a 'text comprehension task'. The cover story that the three tasks were unrelated was enforced by using different fonts and layouts, varying computer tasks with paper and pencil tasks and printing different (fictional) names on the cover sheet for each study.

The first study was the so-called 'word recognition and concentration' study. This study started out by assessing demographics, negative affect and dieting standards in a baseline questionnaire. After finishing the questionnaire, participants completed a lexical decision task in which they were primed with words related to food enjoyment (*hedonic prime and hedonic prime and tell condition*) or neutral words (*neutral prime condition*). Participants were randomly allocated to one of the three priming conditions.

The allegedly 'second' study was introduced as a 'taste test', which, in reality was designed to unobtrusively measure the effect of the priming manipulation on the amount of chocolate consumed. After performing the taste test, the crucial manipulation differentiating the hedonic prime and the hedonic prime and tell condition took place: Participants in the prime and tell condition were told that some of the words in the task they had performed prior to the taste test (the LDT) were related to food enjoyment and that studies have shown that the presentation of these words can influence food intake.

Then, the 'text comprehension study' was introduced. First, negative affect was assessed. Then, participants were asked to highlight key sentences of a scientific article explaining that concentrating on cognitively demanding tasks increases cravings for sugar. The actual purpose of this task was to expose all participants to 'cognitive exhaustion' as a potential explanation for their increased chocolate intake. Participants were then told that the three experiments were now completed, but that we would appreciate it if they could answer some questions to give us feedback about the studies they had participated in. The crucial questions asked participants how difficult and demanding the LDT was to them. These questions encompassed the measure of confabulation.

It was hypothesized that, regardless of their dieting standards, participants in both experimental conditions (hedonic prime condition and hedonic prime and tell condition) would consume more chocolates in the taste test than participants in the control (neutral prime) condition. However, we expected that only participants who had strong dieting goals and who were in the hedonic prime condition (and not in the hedonic prime and tell condition or the neutral prime condition) would use the earlier provided information about the relation between cognitive exhaustion and sugar cravings as a confabulated reason for their higher intake of chocolates.

Materials

Baseline questionnaire. Participants provided demographic characteristics and answered six questions about factors that could affect alertness to enforce the cover story of the first study (i.e., that it investigated the relationship between word recognition and concentration). Baseline hunger and thirst were assessed, and the Positive and Negative Affect Schedule (PANAS) was administered to assess negative affect at baseline

(negative affect T0: α = .74) on scales ranging from (1) *not at all* to (5) *very much*. Dieting standards ('I try to restrict my snack intake') were assessed in a so-called lifestyle questionnaire including 10 filler items (e.g., 'I read the newspaper') that were all rated on scales ranging from (1) *not at all* to (7) *very much*.

Lexical decision task. The LDT in the so-called word recognition and concentration study served to prime participants in the experimental (hedonic prime and hedonic prime and tell) conditions with words related to the concept of food enjoyment and participants in the neutral prime condition with neutral words (Fishbach, Friedman, & Kruglanski, 2003; Stroebe et al., 2008). Each of the 100 randomly presented trials contained a word or a non-word. In total, the task included 50 non-words, 36 neutral words and 14 prime words (hedonic words for the experimental conditions and neutral words for the neutral prime condition), which were each presented for 23 milliseconds, and preceded and followed by a mask (a row of x's; 500 milliseconds). Participants were asked to indicate as fast as possible whether an existing word was presented or not by pressing one of the two buttons on the keyboard.

Taste test. The taste test was introduced as a second, unrelated task. Participants were given 200g of chocolates to taste and were asked to answer several questions regarding various aspects of the chocolates to enforce the cover story that we were interested in participants' ratings of the chocolates. Participants were told to eat as much as they wanted and to take their time as they would need to wait at least 15 minutes before they could move on to the next study. The dependent variable was grams of chocolates consumed. As this variable was skewed, it was transformed using a natural log transformation. However, to facilitate interpretation, means and standard deviations are reported for the non-transformed variable.

Text comprehension task. The 'third experiment' involved filling out the PANAS to assess negative affect post norm-violation (negative affect T1: α = .77). Participants then read a slightly adapted excerpt (1.5 pages) of an article about a paper by Gailliot et al. (2007), which appeared in the APS magazine 'The Observer' (January, 2009). The article posits 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. Participants were told to carefully read the text and to highlight what they thought were the five most important sentences of the article, so that we could assess their comprehension of English texts. The actual purpose of this task was to expose participants to cognitive exhaustion as a potential explanation for eating a lot of chocolate in the taste test.

Confabulation. The measure of confabulation (i.e., the degree to which participants used the information about depletion in the text comprehension task to explain their chocolate intake) involved an evaluation of the LDT in a questionnaire. This questionnaire consisted of seven items that could be answered on 7-point scales ranging from 1 (*totally disagree*) to 7 *(totally agree).*

In the present study (unlike the study by Adriaanse et al., 2014), a factor analysis with varimax rotation indicated that the questionnaire addressed two separate factors, with five items referring to the extent by which the task demanded *Concentration* (e.g., 'The task was too long to really stay concentrated'; α = .83, eigenvalue =3.02) and two items referring to the degree of *Difficulty* of the task (e.g., 'the task was demanding'; $r = .30$, $p < .01$, eigenvalue =1.34). A dieting standards × condition effect was expected for the Concentration subscale but not for the Difficulty subscale of this questionnaire[2] as the text in the comprehension task (see above) specifically stated in the first paragraph that concentration, but not difficulty, increases sugar craving:

> 'It is very important to note, however, that such highly demanding attention tasks are not difficult tasks per se […]. Keeping concentrated during a simple but boring task, is considered to be highly demanding as well'.

Debriefing. At the end of the study, participants were asked what they thought was the purpose of the study. As expected, many participants referred again to the relation between concentration and sugar cravings, and participants in the hedonic prime and tell condition frequently mentioned the presentation of food enjoyment words, which was the explanation that was provided to them. However, none of the participants guessed the actual purpose of the study. All participants were debriefed, thanked and paid.

Results

Randomization check. Separate analyses of variance (ANOVAs) were conducted with condition as the independent variable and dieting standard, baseline negative affect, baseline hunger, baseline thirst, age and BMI as dependent variables. There was a significant difference for age, $p = .03$, but the other effects of condition were not significant, all $ps > .11$. Accordingly, only age was included as a covariate in all of the subsequent analyses.

Manipulation check. An analysis of covariance with condition as between-subjects variable, age as covariate and grams of chocolates consumed as dependent variable was used to determine whether the priming manipulation was successful. As the two prime conditions had up to this point received exactly the same instructions, they were combined into one experimental condition. Condition had a significant

[2]In view of this result, the analyses of the Adriaanse et al. (2014) Study 2 were reran and showed that the effects remain the same when just looking at the concentration subscale.

effect on the amount of chocolates consumed, $F(1,120) = 6.52$, $p = .01$, $\eta_p^2 = .05$. Participants in the experimental conditions ($M = 33.60$, $SD = 25.52$) consumed more than those in the neutral condition ($M = 23.88$, $SD = 23.18$),[3] indicating that the manipulation was successful.

To ensure that dieting standards did not influence the effect of the priming manipulation, a hierarchical linear regression was conducted with the covariate age in the first step, condition in the second step, dieting standards (centred) in a third step and the interaction of condition x dieting standards in the fourth step. Results showed that only the second step, $F_{change}(1, 112) = 4.99$, $p = .03$, including condition, $\beta = .21$, $p = .03$ led to a significant improvement of the model., whereas the subsequent steps including dieting standards, or the interaction term did not ($ps > .25$). This indicates that the effectiveness of the priming manipulation was indeed not affected by dieting standards.

Confabulation: concentration. To determine the effect of condition and the moderator dieting standards on the Concentration subscale of the confabulation questionnaire, two hierarchical regression analyses were conducted. These analyses included confabulation (concentration) as the dependent variable, age as a covariate in the first step, two dummy variables for condition (prime dummy and prime and tell dummy) in the second step and dieting standards (centred) in the third step. Finally, in the fourth step, we tested whether adding the interaction term of the prime dummy × dieting standards or the prime and tell dummy × dieting standards significantly improved the model.

Results indicated that none of the first three steps (all $ps > .69$) or any of the predictors in the first three steps (all $ps > .49$) yielded significant effects. However, in line with expectations, the addition of the prime dummy × dieting standards interaction in the fourth step, $F_{change}(1,109) = 4.80$, $p = .03$, proved to be significant, $\beta = .24$, $p = .03$. A follow-up regression analysis within the hedonic prime condition to decompose the interaction revealed that dieting standards positively predicted confabulation (concentration), $\beta = .36$, $p = .03$. The effect of dieting standards within the other two conditions was not significant, $ps > .37$.

Adding the prime and tell dummy × dieting standards interaction ($p = .52$) or replacing the prime dummy × dieting standards interaction with the prime and tell dummy × dieting standards interaction ($p = .14$) did not significantly improve the model.

Confabulation: difficulty. For the Difficulty subscale, regression analyses yielded no significant effects, all $ps > .23$.

[3]Consumption of chocolates did not differ between the two prime conditions, $p = .603$, which is to be expected since up to this point the manipulation for both conditions was exactly the same.

Negative affect. Negative affect was tested as the proposed underlying mechanism. Two regression analyses were performed with negative affect at T1 as the dependent variable, age and negative affect at T0 in Step 1, two dummy variables for condition (prime dummy and prime and tell dummy) in the second step, dieting standards (centred) in the third step and an interaction term of the prime dummy × dieting standards or the prime and tell dummy × dieting standards (separate analyses) in the fourth step. Unexpectedly, the regression analyses yielded no significant effects, except for a significant effect of negative affect at T0 on negative affect at T1, $p < .001$ in both analyses.

Discussion

Results of Study 1 showed that indeed only those participants with high dieting standards who did not receive an explanation for their increased chocolate consumption scored higher on the confabulation questionnaire. These results provide more convincing evidence for the notion that these participants used the explanation about cognitive exhaustion to explain their indulgent behaviour. In addition, results from this study replicated the moderating role of norm-violation (Adriaanse et al., 2014; Oettingen et al., 2006; Parks-Stamm et al., 2010). No evidence was found for the mediating influence of negative affect.

In the previous studies (e.g., Adriaanse et al., 2014), spontaneous confabulation was assessed rather than directly asking participants to explain their indulgent behaviour to avoid a situation of 'provoked confabulation'. Provoked confabulation refers to instances where people are probed for the reasons for their behaviour (Berlyne, 1972), which increases the likelihood of causal attribution (Wilson, Dunn, Kraft, & Lisle, 1989) and creates a less ecologically valid situation, where participants start searching for reasons for their behaviour regardless of whether the behaviour is norm-violating. Although there are thus good reasons not to directly ask participants to explain their behaviour, a limitation of the more 'spontaneous' assessment of confabulation employed in the current study is that even with stringent control conditions such as the prime and tell condition, we do not have direct evidence that participants were explaining their norm-violating behaviour by reporting that the LDT was very cognitively demanding. Therefore, a study in which participants were explicitly asked for the reasons for their behaviour was added. Although this study does not allow for drawing conclusions about spontaneous confabulations, still, together with Study 1, this should make for particular convincing evidence for the proposition that nonconsciously activated behaviour triggers a tendency to confabulate explanations for this behaviour.

Study 2: Explaining Unexplainable Choices

In this study, participants were unobtrusively primed to prefer either the colour blue or the colour green. We

expected to find that this manipulation would lead to a significantly greater preference for subsequently choosing a cookie in a blue wrapping (from now on referred to as the 'blue cookie') over a cookie in a highly similar green wrapping (from now on referred to as the 'green cookie'), or vice versa (depending on the condition) without participants realizing the effect of the manipulation. Immediately after choosing a cookie, additional information about the cookies appeared below both cookies. After this information was briefly presented, a staged computer error occurred with the intent to blur participants' memory (for details, see Cookie choice task) regarding the timing of this additional information so that we could test whether participants would confabulate by using this post-choice provided information when explaining their choice (cf. Bar-Anan, Wilson, & Hassin, 2010).

Method

Participants. One hundred and eleven students of Utrecht University participated in exchange for 4 euro or course credit. After excluding 14 participants who did not repeat their original choice (see Cookie choice task) and one participant who did not complete the baseline questionnaire, the final sample consisted of 96 students (70 women, 26 men) with a mean age of 21.02 ($SD = 4.03$) and a mean BMI of 21.88 ($SD = 2.83$).

Procedure and design. As a cover story, participants were told that they were participating in two independent experiments, belonging to different experimenters: a colour-classification experiment and a cookie preference task. Different fonts and layouts were used to enhance a sense of difference between the tasks. Participants started by filling out a questionnaire, which included the PANAS to measure baseline levels of negative affect (NA_T0). Participants continued with the colour classification task, in which they were unobtrusively *primed to prefer either the colour blue or green* by pairing one of these colours with negative facial stimuli and the other colour with positive facial stimuli.

Participants then moved on to the cookie preference task and were asked to choose between a blue and a green cookie. Immediately after making the choice, information about the cookies was presented very briefly beneath the cookies. The left cookie was described as 'Triple Choc Light: rich in fibres, low in sugar', and the right cookie was described as: 'Triple Choc Classic: creamy, nice and sweet'. Thus, participants learned after their choice that they had chosen a healthier (Triple Choc Light condition) or a tastier (Triple Choc Classic condition) cookie. We used a staged computer error to make participants believe that this information was provided the entire time (for details, see Cookie choice task).

Participants then filled out the PANAS again to measure negative affect after making the choice (NA_T1) and were subsequently asked why they had chosen their cookie by rating the degree to which different reasons had influenced their decision. Among these reasons were the key reasons that referred to the post-choice provided information: 'because it was the healthiest' and 'because it was the tastiest'.

It was expected that (a) all participants would experience increased negative affect after making their choice and being explicitly asked to explain this choice. This expectation is based on the notion that when people are probed for an explanation that is in reality not accessible (provoked confabulation), this will trigger negative affect and a tendency to make erroneous causal attributions regardless of whether the behaviour is norm-violating or not. That is, whereas people are only expected to spontaneously engage in confabulation when the unconsciously activated behaviour violates a certain standard and thus demands an explanation (e.g., Parks-Stamm et al., 2010), an explicit request to explain the behaviour creates a demand for an explanation by itself, obviating the moderating role of norm-violation.

Moreover, it was expected that (b) participants would partly attribute their choice to the post-choice provided information. In other words, it was expected that participants who learned post-choice that they had selected the Triple Choc Light cookie would confabulate that healthiness had been a more important reason as compared with participants who learned post-choice that they had selected the Triple Choc Classic cookie, and that participants who had learned that they selected the Triple Choc Classic cookie would confabulate that tastiness had been a more important reason as compared with participants who had learned that they had selected the Triple Choc Light cookie. This would be a clear indication of confabulation as this information was not available upon making the choice. Finally (c), it was expected that the measures of negative affect and confabulation would be related. Thus, for participants in the Triple Choc Light condition, negative affect at T1 was expected to be positively correlated with the importance of the reason 'healthiness' and for participants in the Triple Choc Classic condition, negative affect at T1 was expected to be positively correlated with the importance of the reason 'tastiness'.

Materials

Baseline questionnaire. To get a general picture of participants' healthy eating habits, participants were asked whether they tended to eat a lot of sweet or fatty food on a 7-point scale ranging from 1 (*not at all*) to 7 (*very much*). This item was hidden among other filler items (e.g., 'I use soft drugs') in a questionnaire that supposedly assessed factors that could influence reaction times in the colour classification task. To measure negative affect, the PANAS (Watson, Clark & Tellegen, 1988) was administered, using 5-point scales ranging from 1 = *not at all,* to 5 = *very much* (α NA_T0 = .72; α NA_T1 = .77).

Colour classification task. Participants were told that this experiment was designed to test whether people are faster to recognize colours after seeing different types of stimuli, and that they would be randomly assigned to see either words, faces or objects as stimuli right before they would be presented with a colour. In reality, all participants were in the 'faces' condition. Participants were instructed to press either a red, green, yellow or blue button when a matching colour square was presented on the computer screen (500 milliseconds).

Pictures of negative (angry) and neutral or positive (happy) Caucasian faces (derived from the Radboud Faces Database; Langner et al., 2010) were paired with the colour squares (30 milliseconds). A pre- and post-mask depicting a neutral face was applied (400 milliseconds) to ensure that the participants would not become aware of the emotional stimuli and only noticed the neutral faces. The neutral faces were followed by yellow and red squares in both conditions. In one condition, the negative face was followed by a blue square and the positive face by a green square, and in the other condition, the negative face was followed by a green square and the positive face by a blue square. All of the squares (with their accompanying face stimuli) were presented eight times.

Cookie choice task. In this task, two pictures of cookies were shown (derived from Van der Laan, De Ridder, Viergever, & Smeets, 2012) and were displayed in highly similar green and blue packaging.[4] Which cookie (the blue or green cookie) was depicted on which side of the screen was counterbalanced between participants. The participants were told that this was an experiment that was being conducted in collaboration with the Applied Economics Department and that we would like them to select one of the two cookies to taste later on.

Immediately after making their choice, additional information about the cookies was presented very briefly beneath the cookies. One cookie description read: 'Triple Choc Light: rich in fibres, low in sugar', the other read: 'Triple Choc Classic: creamy, nice and sweet'. After 1 second, a staged computer error (Bar-Anan et al., 2010) occurred; the screen turned black, and a picture of a so-called type 404 computer error appeared. The experimenter waited until the participant would report the error and then told the participant that such errors had occurred more often during previous sessions and could be resolved. He then ostensibly 'restarted' the programme. Participants then again saw the two cookies with the additional information. The experimenter then said, 'if you choose the same cookie as you chose earlier, you can continue where you left off'. If participants switched between cookies, they were excluded from data analysis ($N = 14$, see section). The purpose of the staged computer error was to blur participants' memory regarding the sequence of events (i.e., whether the additional information had been present before making the choice) so that we could test whether participants would use this post-choice provided information when explaining their choice (Bar-Anan et al., 2010).

Confabulation questionnaire. After re-choosing their cookie after the staged computer error was 'resolved', participants were told that they would be asked to provide us with the reasons for their choice but that we first would like them to fill out a questionnaire (the PANAS) again. Then, they were indeed asked to rate how much their choice was affected by each of the seven possible reasons ('I chose the cookie, because ….') on scales ranging from 1 (*not at all*) to 7 (*very much*). Four reasons were directly or indirectly related to the colour manipulation (the colour, the package, the general feeling associated with the cookie or the side of the screen), and two reasons ('it was the healthiest' and 'it was the tastiest') related to the post-choice provided information and constituted our measure of confabulation. A last reason stating that the choice was made at random was added for exploratory reasons.

Debriefing. Participants were asked what they thought was the purpose of the experiment after the colour classification task and at the end of the entire study to assess awareness of the manipulations and purpose of the experiment. None of the participants indicated any awareness whatsoever. Finally, participants were debriefed, paid and thanked.

Results

Randomization check. Separate ANOVAs were conducted with condition as the independent variable and age, BMI, baseline negative affect and the tendency to eat sweet and/or fatty foods as dependent variables. A chi-square analysis was conducted to test whether gender was equally distributed across the two conditions. None of the results were significant, all $ps > .11$, indicating that randomization was successful.

Manipulation check. A binominal test showed that there was no overall preference for one colour cookie over the other colour cookie, $p = .26$. A binominal test of the preference for the positive over the negative cookie, however, showed that as expected, cookie choice was indeed successfully manipulated, as 62% of the participants, $p = .03$, chose a cookie with a packaging colour that had been paired with a happy face.

Effects of explanatory vacuum on negative affect. A repeated measures ANOVA, with time as within subject variable, condition as a between subjects variable and negative affect as dependent variable demonstrated a significant effect of time, $F(1, 94) = 11.33$, $p = .001$, $\eta_p^2 = .11$. As expected, participants' negative

[4]A pilot study ($N = 65$) had shown that there was no significant a priori preference for one of the two cookies, $p = .46$.

affect increased from baseline ($M = 1.48$, $SD = 0.38$) to the follow-up measure after creating an explanatory vacuum ($M = 1.61$, $SD = 0.49$). The other effects were insignificant, $ps > .14$.

Effects of explanatory vacuum on confabulation. A Multivariate Analysis of Variance (MANOVA), with condition as a between-subjects variable and the seven reasons as dependent variables, was used to determine whether participants would attribute their choice to the additional information that was provided after the cookie choice. Indeed, participants in the Triple Choc Light condition judged the reason 'it was the healthiest' to be a significantly more important reason ($M = 2.82$, $SD = 1.93$) for their choice than participants in the Triple Choc Classic condition ($M = 1.74$, $SD = 1.11$), $F(1, 94) = 10.96$, $p = .001$, $\eta_p^2 = .10$. None of the other effects were significant (all $ps > .47$) except for a marginally significant effect for the reason 'side of the screen', $p = .08$).

Relationship between negative affect and confabulation. A Pearson's r was computed to assess the relationship between negative affect and the importance of the different reasons for both conditions separately. For participants in the Triple Choc Light condition, negative affect correlated significantly with reason of healthiness, ($r = .32$, $p < .024$). None of the other correlations were significant, all $ps > .15$. For participants in the Triple Choc Classic condition, negative affect correlated significantly ($r = .31$, $p = .035$) with the reason 'I chose the cookie at random' when making the choice. In addition, negative affect correlated marginally significantly with the reason 'It was the tastiest', ($r = .25$, $p = .086$). None of the other correlations were significant, all $ps > .20$.

Discussion

As expected, in both conditions, negative affect increased compared to baseline when participants were asked to provide a reason for their choice. The finding that participants in the control conditions in previous studies (Adriaanse et al., 2014, Studies 1 and 2) reported a decrease in negative affect during the course of the study suggests that the increase in negative affect found in the present study is due to the experience of an explanatory vacuum rather than an unrelated effect of participating in the experiment.

The main aim of Study 2, however, was to conceptually replicate the finding that, in the aftermath of nonconsciously manipulated behaviour, people confabulate an erroneous reason for their behaviour when asked to explain their behaviour. The present study provides a particularly strong demonstration of confabulation, as it shows that even post-choice provided information that cannot possibly have been used during the selection process retrospectively influenced what participants reported as the reasons for their choice. That is, participants who, after choosing their cookie, learned that they chose a healthier cookie deemed healthiness as a more important reason for their choice than those who had chosen a tastier, less healthy cookie. We did not observe a significant effect of condition for tastiness. One reason for this may be that the additional information provided post-choice may have more clearly related to the dimension of healthiness than to the dimension of tastiness.

Importantly, and in line with recent findings demonstrating the mediating role of negative affect (Adriaanse et al., 2014), for participants in the Triple Choc Light condition, negative affect at follow-up was significantly positively related to the degree to which they attributed their choice to the healthiness of the cookie. For participants in the Triple Choc Classic condition, negative affect at follow-up was significantly positively related to the degree to which these participants reported having made their choice at random and only marginally significantly positively related to the reason of tastiness. A possible explanation for the relation to the reason of randomness could be that, in particular, for people concerned with their weight, tastiness is less of a self-promoting reason than confabulating that the choice was made at random.

General Discussion

The present studies were designed to provide further evidence for confabulation as a downstream consequence of nonconsciously activated behaviour. Results of Study 1 demonstrated that of the participants primed to indulge in chocolate, only those who had high dieting standards and who did not receive an explanation for their increased chocolate consumption rated the LDT they performed prior to consuming chocolates as more cognitively exhausting. In other words, higher scores on the confabulation questionnaire were found only in participants for whom indulging in chocolate was considered norm-violating and who experienced an explanatory vacuum and not as a result of norm violation in general. This provides more convincing evidence for the suggestion that results were indeed indicative of attempts to explain the norm-violating behaviour rather than, for example, representing a mood congruency bias. In Study 2, a different approach was used to test the notion that in the aftermath of nonconsciously manipulated behaviour, people confabulate a reason for their behaviour. In this study, rather than inferring confabulation, participants were directly probed for the reasons of a nonconsciously influenced choice. Adopting a staged computer error paradigm (cf. Bar-Anan et al., 2010), it was found that post-choice provided information influenced participants' ratings of the reasons for their choice. Taken together, results from Studies 1 and 2 provided additional support for the assumption that as a result of nonconsciously activated (i.e., unexplainable) norm-violating behaviour people resort to confabulating plausible reasons for their behaviour. In addition, Study 1 replicated previous studies demonstrating the moderating role of

norm-violating behaviour on non-provoked confabulation (Adriaanse et al., 2014; Parks-Stamm et al., 2010).

The present findings provide further evidence for the suggestion that the influence of nonconscious processes on behaviour may extend beyond executing the primed behaviour. In view of the significance of nonconscious processes in our daily lives and the large number of studies that have been conducted on this topic, it is remarkable how little attention has been given to the psychological after-effects of nonconscious goal striving. This is particularly surprising when thinking about nonconscious goal striving in the context of classic findings such as research on the limits of introspection (Nisbett & Wilson, 1977) and research on cognitive dissonance theory (e.g., Stone & Cooper, 2001), which clearly point to the potential for post hoc attribution processes in this context (Bar-Anan et al., 2010). In the past decades, research has mainly focused on demonstrating that nonconscious behaviour instigation occurs in different areas and on understanding its underlying processes (e.g., Hassin, Uleman, & Bargh, 2005) and, in case of unwanted behaviours, how these nonconscious influences can be overruled. However, the psychological 'after-effects' of nonconsciously activated behaviours, indicating how people deal with these behaviours, have received little attention. The present findings add to the emerging literature to this topic suggessing that there are relevant psychological downstream consequences that warrant further investigation, for example, on their long-term effects (see below).

The present findings also add to the literature on cognitive dissonance theory. That is, the present studies show that the assumption that people are motivated to restore the inconsistency between two cognitions also applies to nonconsciously activated behaviour that is inconsistent with conscious standards. Thus, the present research aligns with work on Cognitive Dissonance Theory by suggesting that inconsistency between two cognitions motivates attempts to justify the behaviour and restore consistency. However, the present line of reasoning also departs from Cognitive Dissonance Theory by suggesting a different 'route of justification' (i.e., confabulating reasons for the behaviour vs. classical dissonance reduction through attitude change) to solve the inconsistency (Adriaanse et al., 2014). In addition, it should be noted that unlike in the present studies, in dissonance studies, participants are never completely unaware of the cause of their behaviour (usually some nudging by the experimenter that is followed voluntarily). 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 experience a complete lack of justification for their behaviour. In this sense, one could argue that nonconsciously activated behaviours provide a particular suitable and plausible context for feelings of cognitive dissonance, making it even more surprising that these processes have hardly been investigated in the aftermath of nonconsciously activated behaviours (cf. Bar-Anan et al., 2010).

The present findings may also have important practical implications. For example, while this is not tested in the present paper, it makes sense to assume that there may be potential long-term effects of confabulations. To illustrate, consider a female dieter who is primed to indulge in chocolate and who blames a negative event that happened earlier that day for violating her diet. Having adopted this explanation will increase the accessibility and plausibility of this explanation of her behaviour and thus the chance of using this as a confabulated reason for future behaviours (Nisbett & Wilson, 1977). In line with research demonstrating that people incorporate misattributed internal states into their self-concept and act consistently on them (e.g., Bem, 1972; Fazio, Effrein, & Falender, 1981; Swann, Rentfrow, & Guinn, 2002), it is not unlikely that this dieter may eventually view herself as an emotional eater and even start acting accordingly. Bar-Anan et al. (2010) already provided some first evidence that in the case of provoked confabulation, confabulated reasons become integrated into self-knowledge and affect subsequent behaviours. Future research is necessary to further understand the potential long-term effects of confabulation processes.

The present findings also trigger the question of whether confabulation after nonconscious behaviour activation is an adaptive or maladaptive process. Although this question cannot be answered in the present paper and should be scrutinized in future studies, there are also reasons to believe that in general, this process is likely to be adaptive (Chartrand et al., 2010). First of all, previous research on attribution has demonstrated that people are inclined to attribute their behaviour to reasons that are self-promoting (Kunda, 1990). Moreover, considering that people have a strong need to maintain a positive standing regarding their identity goals (e.g., being a healthy eater), or in other words that the 'self is a self-symbolizing animal' (Wicklund & Gollwitzer, 1982) and that people are inclined to attribute success to internal reasons and failure to external ones beyond their control (Mezulis, Abramson, Hyde, & Hankin, 2004), it is likely that confabulation can serve a self-promoting function. However, considering the example above where confabulation leads to a self-fulfilling prophecy of someone being an emotional eater, maladaptive outcomes should be considered as well.

Several limitations of the present studies should be noted. First of all, Study 1 yielded no evidence for the proposed underlying mechanism of negative affect although there was a trend in the expected direction. As previous research demonstrated that confabulation occurs reflexively upon encountering an explanatory vacuum, a plausible explanation for this unexpected finding could be that at this point, negative affect was already attenuated through confabulation (Parks-Stamm et al., 2010). Study 2 did yield the expected increase in negative affect, although it should be noted that this study did not include a control condition. A second limitation is that although Study 1 was designed

to test spontaneous confabulation, we should be careful in using this term. That is, although in this study participants were never probed for an explanation, still the opportunity to confabulate was provided to them.

Notwithstanding these limitations, the present studies provide more convincing evidence for the phenomenon of confabulation as a downstream consequence of nonconsciously activated behaviour. The present line of research contributes to the literature on nonconscious processes and helps to explain how people explain the unexplainable.

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

Adriaanse, M. A., Weijers, J., De Ridder, D. T. D., De Witt Huberts, J., & Evers, C. (2014). Confabulating reasons for behaving bad: The psychological consequences of unconsciously activated behaviour that violates one's standards. *European Journal of Social Psychology*, *44*, 255–266.

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.

Bem, D. J. (1972). Self-Perception Theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (pp. 1–62), 6. New York: Academic Press.

Berlyne, N. (1972). Confabulation. *British Journal of Psychiatry*, *120*, 31–39.

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.

Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of Personality and Social Psychology*, *41*, 232–242.

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

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.

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.) (2005). *The new unconscious*. New York: Oxford University Press.

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

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

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, *24*(8), 1377–1388.

Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*, 711–747.

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.

Sheeran, P., Aarts, H., Custers, R., Rivis, A., Webb, T. L., & Cooke, R. (2005). The goal- dependent automaticity of drinking habits. *British Journal of Social Psychology*, *44*, 47–63.

Sheeran, P., Gollwitzer, P. M., & Bargh, J. A. (2013). Nonconscious processes and health. *Health Psychology*, *32*, 460–473.

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.

Swann, W. B. Jr., Rentfrow, P. J., & Guinn, J. (2002). Self-verification: The search for coherence. In M. Leary, & J. Tagney (Eds.), *Handbook of self and identity* (pp. 367–383). New York: Guilford.

Van der Laan, L. N., De Ridder, D. T. D., Viergever, M. A., & Smeets, P. A. M. (2012). Appearance matters: Neural correlates of food choice and packaging aesthetics. *PloS One*, *7*(7), e41738.

Wansink, B., & Sobal, J. (2007). Mindless eating: The 200 daily food decisions we overlook. *Environment and Behaviour*, *39*, 106–123.

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*, *47*, 1063–1070.

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

Wicklund, R. A., & Gollwitzer, P. M. (1982). *Symbolic self-completion*. Hillsdale, N.J: Lawrence Erlbaum.

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* (pp. 287–343), 22. Orlando, FL: Academic Press.