Chapter 12

When Consciousness Needs to Explain Unconsciously Activated Behavior

Ana P. Gantman[1], Peter M. Gollwitzer[2,3], and Gabriele Oettingen[2,4]

[1] Department of Psychology, Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA

[2] Department of Psychology, New York University, New York, NY, USA

[3] Department of Psychology, University of Konstanz, Germany

[4] Department of Psychology, University of Hamburg, Germany

Abstract: Goals can be activated outside of awareness, such that goal-directed behaviors can be triggered and deployed outside of conscious awareness. In the case of nonconsciously activated behavior, if the outcome violates personal standard or norms, recognizing that there is no explanation for the outcome triggers negative affect (i.e., the explanatory vacuum arises). We review research aimed at understanding the explanatory vacuum, including new work aimed at investigating the role of perceived agency and control. While the results of these latter experiments were not straightforward, we will discuss them in detail in the spirit of transparency and in the hope that they spark novel avenues for future research. In particular, we try to be responsive to Julius Kuhl's appeal that research on self-regulation should focus on the interaction of different self-regulatory systems (in our case the interaction of consciousness and nonconsciously activated behavior), and generate related novel research ideas.

Research on goals has become more and more accepting of the possibility that goals activated outside of a person's conscious awareness are also quite powerful determinants of people's thought, feelings, and actions (e.g., Bargh, Schwader, Hailey, Dyer, & Boothby, 2012; Bargh, Gollwitzer, & Oettingen, 2010; Hassin, Uleman, & Bargh, 2004). Research, however, is just beginning to investigate how consciousness deals with the fact that nonconscious goal striving produces goal-related outcomes as well (Oettingen, Grant, Smith, Skinner, & Gollwitzer, 2006; Parks-Stamm, Oettingen, & Gollwitzer, 2010; Bar-Anan, Wilson, & Hassin, 2010; Adriaanse, Weijers, De Ridder, De Witt Huberts, & Evers, 2014). An exception is Julius Kuhl's research on his personality systems interactions (PSI) theory (Kuhl, 2000; see also Chapter 2 in this volume). PSI theory makes the important point that the outcome of a person's self-regulation of action (i.e., her goal striving) is best understood in terms of the coaction of explicit and implicit processes pertaining to affect and

cognition, and that these processes – as well as their coaction – should be moderated by the action-control relevant personality attributes of state vs. action orientation. This theoretical perspective has produced many provocative theoretical insights and empirical findings. For instance, based on a review of the empirical literature, Kuhl and colleagues inferred that the common finding that religious individuals show higher levels of emotional well-being (even though they must frequently endure aversive tasks and forsake pleasurable experiences) is based on an implicit rather than an explicit mode of self-regulation; whereas the implicit mode is quite automatic in the sense of being efficient and operating outside of awareness, the explicit mode is characterized by a conscious effort to guide one's thoughts, feelings, and actions (Koole, McCullough, Kuhl, & Roelofsma, 2010). In the present chapter, we will address the question of what will happen when people's implicit and explicit modes of self-regulation encounter each other. More specifically, we will discuss how people consciously process the outcomes produced by their nonconscious goal striving; we have referred to this challenge as dealing with an explanatory vacuum (Oettingen et al., 2006). By addressing this issue we try to be responsive to Julius Kuhl's appeal that research on self-regulation should focus on the interaction of different self-regulatory systems (in our case the interaction of consciousness and nonconsciously activated behavior).

Behavior Activated Outside of Awareness

Desired end state representations (i.e., goals; Dijksterhuis & Aarts, 2010) can be activated outside of the awareness of the actor, such that goal-directed behaviors can be triggered and deployed outside of conscious awareness (Hassin, Uleman, & Bargh, 2004). According to Auto-Motive Theory, goals may be activated outside of awareness through the repeated pairing of a given situation and its related goal; the contextual cues eventually activate the goal through the established associative link (Bargh, 1990; Bargh & Gollwitzer, 1994). This model predicts that both conscious and nonconscious activation of goals should lead to similar goal attainment rates and qualities of goal striving (Bargh, Gollwitzer, Lee-Chai, Barndollar, & Trötschel, 2001; Chartrand & Bargh, 2002). Indeed, nonconsciously activated goals exhibit hallmarks of goal pursuit. In particular, nonconscious goals, like conscious goals, lead to goal-directed action, stay active until completed, produce persistence in the face of set-backs, and resumption after interruption (Bargh et al., 2001). Nonconscious goal triggers in the environment can include the verbal description or the presence of desired outcomes (Gollwitzer, Sheeran, Trötschel, & Webb, 2011), the presence of a significant other (Fitzsimons & Bargh, 2003), means that are often used to attain a goal (Shah & Kruglanski, 2003), or temptations that frequently interfere with goal pursuit (Fishbach, Friedman, & Kruglanski, 2003). The goal is activated outside of awareness, but the goal striving itself (i.e., the behaviors that serve to attain the goal) can be consciously engaged in even when the source is not consciously accessed.

Indeed, individuals may be unaware of one or more aspects of the processes that underlie the direct guidance of behavior by the environment. They may be unaware of (1) the environmental cues triggering the behavior (e.g., it may be presented below the threshold for awareness), (2) the link between the environment and the behavior (e.g., the agent is unaware of the relationship between the environmental cue and the behavior), or (3) the outcome of that process (e.g., behavioral mimicry; Chartrand, 2005). Because the agent does not know about one or more of these aspects, they can be said to be "introspectively blank" – if asked for an explanation for the behavior, they cannot provide (a veridical) one (for a review, Chartrand, 2005). Critically, as we will review below, the introspective blank may be regarded negatively by the individual, and can be reflexively filled by confabulation.

In the case of nonconsciously activated behavior, most often, people are aware of the outcome (e.g., behavior, decision, preference) but unaware of either the environmental cue or its relation to the behavior. In some cases, if the outcome violates personal standards or norms, recognizing that there is no apparent explanation for the outcome triggers negative affect. This phenomenon is referred to as the explanatory vacuum.

What Is the Explanatory Vacuum?

The full model of the explanatory vacuum phenomenon (Adriaanse et al., 2014; Oettingen et al., 2006) begins with a nonconsciously activated behavior. When the behavior violates expectations, negative affect arises, and individuals devise an explanation for their behavior (i.e., confabulate). In other words, the two consequences of recognizing unexpected, nonconsciously activated behavior are negative affect and confabulation. To better understand the phenomenon, we will discuss each consequence in turn.

Negative Affect

For this chapter, we are using the term negative affect broadly to include dissonance, uncertainty, arousal, and others (see Proulx, Inzlicht, & Harmon-Jones, 2012). The first study to investigate the affective consequences of acting in an explanatory vacuum operationalized negative affect using the positive and negative affect scale (PANAS; Oettingen et al., 2006, Study 1). In this experiment, all participants engaged in a task that required cooperation. Prior to the task, they were given either a conscious or a nonconscious goal to be competitive (i.e., norm-violating) or cooperative (i.e., norm-conforming). While the affective consequences of conscious versus nonconscious goal pursuit have been found to be similar in previous work – success (failure) in both conscious and nonconscious goal pursuit may lead to a positive (negative) mood (Chartrand & Bargh, 2002) – when successful goal pursuit involves behavior that violates a norm, only nonconscious (vs. conscious) goal striving lacks an apparent explanation (Oettingen et al., 2006). With no explanation for the nonconsciously activated, norm-violating competitive behavior, recognizing this "introspective blank" elicits negative affect.

Accordingly, participants given the nonconscious goal of competitiveness showed heightened negative affect as a result of their conflicting, norm-violating behavior (namely, acting competitively in a cooperation-based task) relative to those with a conscious goal to act competitively, and those with either a conscious or nonconscious goal to cooperate. Only participants in the nonconscious norm-violating condition experienced inconsistency between their norms and their behavior without apparent cause. We suggest that they could not explain their behavior led them to experience increased negative affect. Nonconsciously activated behavior triggering negative affect in a similar manner was recently replicated in the domains of antisocial and eating behavior (Adriaanse et al., 2014). Critically, however, the precise nature of the negative affect is unknown.

It is difficult to determine whether the negative affect derives from only the particular unexplained negative behavior or from the lack of an explanation more generally. In other words, participants may feel negative affect because they recognize that they have committed a negative behavior (e.g., act competitively in a cooperative context). But it is also possible that the negative affect arises because recognizing unexpected behavior (i.e., the introspective blank) threatens personal control and agency. In the former case, confabulation would serve to understand the origin of the specific

action. In the latter case, we would expect confabulation to be primarily aimed at restoring personal agency and control. To better understand negative affect associated with an explanatory vacuum, and the explanatory vacuum more generally, we turn to the issue of confabulation.

Confabulation

Confabulation is a term whose origins lie in clinical psychology. Confabulation was originally considered a disorder of memory, as patients with Korsakoff syndrome with severe amnesia would report as memories events that either did not happen or had happened much earlier in the patient's life (Hirstein, 2005). The definition of confabulation expanded to include denials of ailments, known as anosognosia, misidentification syndromes such as Cap-gras syndrome, and the explanations of corpus callosotomy patients regarding behavior derived from linguistically inaccessible content. Critically, such confabulations are genuinely believed by patients and delivered with conviction – with no intent to deceive. Dennett writes, "These strings or streams of narrative issue forth as if from a single source – not just in the obvious physical sense of flowing from just one mouth, or one pencil or pen, but in a more subtle sense: their effect on any audience is to encourage them to (try to) posit a unified agent whose words they are, about whom they are: in short, to posit a center of narrative gravity" (1993, p. 418). We know that such behavior is not limited to clinical samples – people unknowingly attribute their actions to the wrong causes (Bem, 1967, Nisbett & Wilson, 1977, Wheatley, 2009) – and in both cases, it appears that the motivation is partly to restore a sense of "agentic coherence," the idea that the actions or beliefs make sense coming from that particular person, to reflexively form causal theories to explain events (Cooney & Gazzaniga, 2003).

In the confabulation literature, there is a distinction made between those formed reflexively – spontaneous confabulations – and those given in response to a question by an authority figure, known as provoked confabulations (Kopelman, 1987). We can find evidence for provoked confabulation in healthy individuals (Nisbett & Wilson, 1977). When people are asked to explain their behavior, rather than admitting that they do not know where their action or preference came from, they willingly provide an explanation anyway. For example, participants were presented with an array of stockings as if they were in a consumer study and asked to select their preference. Participants chose the rightmost pair despite all pairs being identical. When asked the reason for their choice, participants did not mention the position of the stockings and some even refuted this as a possibility (Nisbett & Wilson, 1977). When asked directly about the possibility of a position effect, participants denied it and felt "either that they had misunderstood the question or were dealing with a madman" (Nisbett & Wilson, 1977, p. 244).

Participants have also been shown to confabulate about snack preferences. Participants were seated in a room with a confederate who ate either goldfish or animal crackers. Participants mimicked the eating behavior of the confederate, such that participants with an animal cracker-eating confederate ate more animal crackers than goldfish crackers whereas participants with a goldfish-eating confederate ate more goldfish than animal crackers. Critically, participants were not aware that they were mimicking the experimenter and when asked, reported a preference for the cracker that they ate more of. In this study, mimicry mediated the relationship between eating behavior and cracker preference (Tanner, Ferraro, Chartrand, Bettman, & Baaren, 2008).

Bar-Anan and colleagues (2010) have argued that post-priming misattribution may be a common, everyday example of confabulation. They have shown that when male participants are primed with romantic goals, they will choose a course given by a female (vs. male) instructor, regardless of the course's actual topic; yet, when asked, participants reported that the course's topic was the primary reason for their choice. In another study, participants primed with a goal to earn money were more likely to prefer a game with pictures of American presidents as they appear on American money,

over another game that depicted normal pictures of the same American presidents compared to those with a neutral prime. It was only after indicating their preference for either one of these games that participants received information about the games' difficulty. 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. When asked, participants misattributed a choice to a cue that was actually provided after they had already made the choice. In the case of confabulation, we see evidence for a desire to have performed explicable behaviors, yet it is still difficult to disentangle whether this implicates a desire to create a "center of narrative gravity" (Dennett, 1993, p. 418) or whether the situational norms demand a reasonable explanation for performing a given behavior.

Negative Affect Partially Mediates Confabulation

According to Parks-Stamm et al. (2010) confabulation is likely a consequence of the negative affect triggered by acting in an explanatory vacuum. Because the negative affect is an aversive state that people are motivated to reduce, participants acting in an explanatory vacuum confabulate an alternative explanation for their behavior. In an attempt to test this proposed sequence of events, Adriaanse and colleagues (2014) analyzed whether the tendency to confabulate was indeed mediated by negative affect. In one study, participants played a video game that primed either neutral or antisocial content. Next participants were asked to complete an ostensibly unrelated task for which they were told they would receive no credit. They were asked to help a fellow student with a tedious computer task and to stop when they felt they had sufficiently helped. Then they filled out an exit survey about the new lab space they had taken the study in. Participants primed with antisocial behavior completed fewer help trials, experienced greater negative affect, and provided a more negative evaluation of the lab space (e.g., the chair was uncomfortable) than participants primed with neutral content. As hypothesized, the negative affect experienced after the primed antisocial behavior mediated the relationship between priming condition and lab space evaluation (i.e., confabulation). In other words, unexpected antisocial behavior led to increased negative affect, which in turn, drove participants to confabulate an explanation for their behavior (e.g., I stopped helping because the chair was uncomfortable). These findings echo classic work distinguishing cognitive dissonance from self-perception via the role of arousal (Elliot & Devine, 1994; for more on the role of negative affect in cognitive dissonance see also Chapter 11 in this volume). In summary, recognizing unexpected behavior (i.e., the explanatory vacuum) leads to feelings of negative affect, which triggers confabulation aimed at explaining the behavior. Again, it is not yet clear whether the confabulation processes point to a desire to explain only the negative behavior or the notion of enacting unexplained behavior more generally.

Moderators of the Explanatory Vacuum

Two of the potential moderators of the explanatory vacuum phenomenon offer mixed evidence for understanding the nature of the explanatory vacuum. The explanatory vacuum phenomenon is comparatively more likely when the behavior demands an explanation, meaning that it has violated personal standards or social norms. This suggests that the negative affect may derive from needing an explanation for that specific behavior, and may not evoke concepts from theories of conscious will. Yet, the explanatory vacuum is also more likely when a prior explanation is not available. This coincides with the need for a prior consistent thought to precede the action to experience the feeling

of conscious will (Wegner, 2002). In other words, the prior consistent thought obviates the negative affect associated with the explanatory vacuum.

The Behavior Demands an Explanation

Not all nonconsciously activated behavior triggers negative affect and a subsequent need to explain the behavior. Negative affect and a tendency to confabulate only arise when the nonconscious behavior *demands* an explanation (e.g., because it violates a norm or a consciously held personal standard; Oettingen et al., 2006; Parks-Stamm et al., 2010). The moderating role of norms or standards was demonstrated in the aforementioned study by Oettingen and colleagues (2006), who reported increased negative affect only in the nonconscious, norm-violating condition, but not in the nonconscious norm-conforming condition. This moderating role of standards on negative affect was replicated and extended, demonstrating participants' tendency to confabulate in the health domain (Adriaanse et al., 2014). Participants with either high or low dieting standards completed a lexical decision task to prime them with neutral or hedonic words. Then, after engaging in a subsequent taste-test, which unobtrusively measured chocolate intake, confabulation was assessed by measuring to what extent participants, after reading a text proposing that cognitively demanding tasks increase cravings for sugar, retrospectively reported that the lexical decision task that had preceded the taste test was cognitively exhausting. Evidence for a mediated moderation model was obtained, suggesting that an interaction between dieting standards and the priming condition (i.e., having high dieting standards and being primed with a hedonic orientation) led to increased confabulation regarding how cognitively exhausting (and thereby chocolate consumption-justifying) the lexical decision task was. Experiencing negative affect partially mediated the interaction effect of priming and personal standards on confabulation; when the prime conflicted with personal standards, negative affect triggered confabulation (Adriaanse et al., 2014; Study 2).

Amending the studies by Oettingen et al. (2006) and Adriaanse et al. (2014), Bar-Anan et al. (2010) demonstrated that "provoked confabulation" (Berlyne, 1972) may follow nonconsciously activated behaviors that were not norm-violating. That is, evidence for post-priming confabulation was found for neutral behaviors when participants were explicitly asked to explain their behavior. So, rather than norm-violation evoking the need for explaining the behavior, in this latter case it is simply the explicit request to explain one's choices or actions that leads to the "introspective blank" that triggers confabulation. We suggest that it is only in the case of non-provoked or spontaneous confabulation that norms or standards act as a moderator. In the case of an explicit request to explain one's behavior, merely asking for an explanation is likely to be sufficient to trigger confabulation regardless of whether the behavior is norm violating or not.

When No Other Explanation Is Available

There is also evidence that the explanatory vacuum leads to spontaneous confabulation (Parks-Stamm et al., 2010), much in the way that the dissonance literature demonstrates that counter-attitudinal behaviors are *reflexively* interpreted as indicative of one's own attitudes (Festinger, 1962; Lieberman, Ochsner, Gilbert, & Schacter, 2001). Parks-Stamm et al. (2010) hypothesized that the increased negative affect in the nonconscious goal condition arose specifically from the lack of explanation for the behavior. The authors found that the heightened negative affect in the nonconscious goal condition could be reduced when a plausible explanation for primed competitive behavior was made available. More specifically, participants were given a cooperative task to complete with a partner, in which acting quickly was synonymous with acting competitively. Prior to this col-

laborative task, participants were asked to perform a seemingly unrelated task which half of the participants had to perform quickly and the other half accurately. Of the participants in the explanatory vacuum (e.g., enacting a nonconscious competitive goal in a cooperative task), those who engaged in the prior speed task showed less negative affect than those in the accuracy task. When primed goal-directed behaviors can be explained (i.e., by having just done a task as quickly as possible), this obviates the negative affect associated with the explanatory vacuum. Importantly, it made no difference whether participants were asked to reflect on their performance in the cooperative task, suggesting that the prior goal was automatically taken up as an explanation for norm-violating behavior.

Critically, norm-violation *per se* does not seem to be the cause of the increase in negative affect and the tendency to search for explanations. Negative affect did not arise when the norm-violation was the result of a consciously held, norm-violating goal (Oettingen et al., 2006). Similarly, when participants are told about the priming condition, they do not experience negative affect or confabulate (Adriaanse, Kroese, Weijers, Gollwitzer, & Oettingen, 2016). Similar to Adriaanse et al. (2014, Study 2) participants with either high or low dieting standards were included and completed a lexical decision task to prime them with neutral or hedonic words. After engaging in the taste-test, participants in the hedonic prime-and-tell condition (but not in the regular hedonic prime condition or in the neutral condition) were told that the lexical decision task may have affected their food intake. Results showed that participants with high dieting standards in the hedonic prime condition, but not in the hedonic prime-and-tell condition, used the explanation about cognitive exhaustion as a confabulated reason to explain their indulgent behavior. Apparently, confabulation arises as a psychological consequence of acting in an explanatory vacuum and not of norm-violation in general.

Taken together, this suggests that when goal-primed behavior demands an explanation, this is aversive, and so people provide their own explanation even at the expense of accuracy. This leaves open the possibility that the experience of recognizing the "introspective blank" is aversive because it threatens one's sense of agency and personal control. We do not yet know however, whether the negative affect associated with the explanatory vacuum derives from concern about performing that *specific* behavior or whether the recognition of unexpected actions threatens a sense of perceived agency or control over one's actions *generally*. In other words, the explanatory vacuum phenomenon points to one of two possibilities: Either the explanatory vacuum highlights a norm of explanation for a critical behavior or a motivation for agency or "agentic coherence."

Explanatory Vacuum and Cognitive Dissonance

There are some obvious parallels between the affective consequences of acting in an explanatory vacuum and classic work on cognitive dissonance (Elliot & Devine, 1994; Festinger, 1962; Stone & Cooper, 2001; Harmon-Jones, Harmon-Jones, Amodio, & Gable, 2011) which has shown that people experience discomfort (dissonance) when an inconsistency exists between a person's behavior and her respective attitudes. However, whereas a typical dissonance study creates a situation of insufficient justification for the behavior (usually a pressing request by the experimenter to behave in a way that is inconsistent with one's attitudes), acting in an explanatory vacuum entails a situation of no justification. Accordingly, in this latter, more extreme case of limited justification, discomfort is experienced (Adriaanse et al., 2014). Moreover, the explanatory vacuum differs interestingly for the action-based model of cognitive dissonance (see Chapter 11 in this volume), as the negative affect occurs after an action as been performed. Moreoever, the evaluation of the performed action is likely relatively negative, vis-a-vis when evaluation of a chosen action is bolstered by the spreading of decision alternatives. Still, there are similarities to acknowledge. For instance, cognitive disso-

nance is most likely under conditions of high (vs. low) choice (Brehm, 1956), and the experience of nonconscious goal pursuit likely resembles high choice, as there is no conscious access to additional factors influencing the choice outside of one's control. Some (Adriaanse et al., 2014; Bar-Anan et al., 2010; Parks-Stamm et al., 2010) have proposed that – very much like the misattribution to attitudes in cognitive dissonance paradigms – another psychological consequence of nonconsciously activated behavior is a tendency to misattribute the behavior to erroneous causes. For this paper, we consider this form of misattribution as a kind of confabulation.

In our lab, we have attempted to understand the role of agency and control in the explanatory vacuum phenomenon. One way to think about the potential role for agency is through the lens of dissonance. The reasoning would be as follows: one attitude that people have is that they are the general cause of their actions (Wegner, 2002) and being presented with one's own nonconsciously activated behavior constitutes a counter-attitudinal behavior and so negative affect (i.e., dissonance) arises and confabulation serves to recover the attitude that one is in control of one's actions. This reasoning does not quite work as the explanatory vacuum phenomenon is not a form of dissonance but hopefully probes the desired intuiton about how nonconsciously activated behavior relates to agency. As previously mentioned, dissonance occurs under conditions of high choice, and of course, participants do not choose when an environmental cue triggers behavior outside of awareness. This is worth outlining however, because the general intuition is the same. We reasoned that recognizing behavior that cannot be explained could threaten perceived personal agency and control. Usually, when people feel that they are the author of their actions, they perceive the thought of the action as having preceded that action (Wegner, 2002). In the case of nonconscious goal pursuit, and in particular the explanatory vacuum, there is no awareness of the thought, and without it, it seems that there is little room for the feeling of agency.

Explanatory Vacuum and Perceived Agency and Control

We conducted a series of experiments aimed at understanding the relationship between the explanatory vacuum and perceived agency and control. These experiments were inconclusive, but we will discuss a few of them in detail in the spirit of transparency and in the hope that they spark novel avenues for future research on this topic (all materials available upon request).

We hypothesized that participants would experience negative affect after recognizing a neutral/positive nonconsciously activated behavior. We devised a novel paradigm for investigating the explanatory vacuum that utilized a single neutral/positive behavior (disclosure) and modified context to render more or less of that behavior unexpected. This constituted the first step in going beyond previous research on the explanatory vacuum to determine whether the explanatory vacuum could reveal a threat to personal control and agency. We did see some evidence that recognizing neutral/positive nonconsciously activated behavior could lead to negative affect. The first experiment was designed to test whether the explanatory vacuum phenomenon generalizes beyond acting on a negative behavior. Previous research has identified the explanatory vacuum phenomenon in association with negative behaviors (e.g., acting competitively in a cooperative context). If the phenomenon does not extend beyond negative behaviors, the negative affect is likely a consequence of the specific unexplained negative behavior.

To test this question, we created a novel paradigm for studying the explanatory vacuum phenomenon. Previous studies have primed negative behavior (e.g., competitive behavior, snacking behavior), and so it cannot yet be discerned whether the arousal associated with the explanatory vacuum is associated only with the execution of specific, undesired behaviors (i.e., competitiveness) or the

more general phenomenon of a desire for perceived personal agency. We chose the act of disclosure, a neutral (or sometimes positive; Tamir & Michell, 2012) behavior, and changed whether or not the act of disclosure would be the expected behavior in the situation by manipulating the context. Specifically, participants were either primed with disclosure or non-disclosure, and the anticipated recipient of the personal information was trustworthy or not. This effectively manipulated whether it makes sense to disclose to them (i.e., because they are trustworthy with personal information) or not. This created two "mismatch" or explanatory vacuum conditions (i.e., disclosing to an untrustworthy experimenter, and failing to disclose to a trustworthy experimenter) and two "matched" or control conditions (i.e., disclosing to a trustworthy experimenter, and failing to disclose to an untrustworthy experimenter). Self-reported affect was our main dependent variable of interest, as the first experiment sought to investigate whether negative affect follows the recognition of neutral/positive, unexpected, nonconsciously activated behavior. We also measured possible avenues for confabulation as assessed in previous explanatory vacuum studies (see Adriaanse et al., 2014). Surprisingly, we found that participants who received the nondisclosure priming and interacted with the trustworthy experimenter experienced the greatest negative affect, suggesting that perhaps the norms of compliance in the study were strong and so failing to comply constituted norm-violating or antisocial behavior.

We did not expect that participants who received nondisclosure priming and a trustworthy experimenter would experience relatively more negative affect than those who received the disclosure priming and the untrustworthy experimenter. We simply did not foresee this, as we thought that participants in the disclosure condition and an untrustworthy experimenter would be more likely to recognize and be surprised by their behavior. Against our predictions, however, there was no observable difference between participants who were primed with disclosure and had either a trustworthy or untrustworthy experimenter.

Trying to further understand the negative affect associated with the explanatory vacuum, we employed a misattribution of arousal paradigm in another study in order to manipulate whether participants had an explanation for their negative affect. Previous research has found that when participants were able to attribute their feelings of dissonance to a pill (described as causing feelings of discomfort), they no longer showed evidence of attitude change normally associated with dissonance (Zanna & Cooper, 1974). Similarly, we reasoned that if the negative affect associated with the explanatory vacuum was due to threat to personal control and agency, then participants with no explanation for their negative affect would report greater desire for control (to compensate for the threat; Burger & Cooper, 1979) than those who were given an explanation.

All participants were assigned to one of the two explanatory vacuum conditions (i.e., disclosure priming and untrustworthy experimenter, and nondisclosure priming and trustworthy experimenter). We also manipulated the presence of a plausible explanation for negative affect: All participants were exposed to unpleasant construction noise, and were either given an explanation for their experienced affect (i.e., a misattribution stimulus) or not.

This study also yielded mixed results. In the disclosure explanatory vacuum, participants reported decreased desire for control. In the nondisclosure explanatory vacuum, participants reported increased desire for control as we had hypothesized. While the need for control is implicated in the explanatory vacuum phenomenon, the two conditions do not affect desire for control in the same way. For participants in the disclosure explanatory vacuum, it appears that the negative affect they experience is related to the specific action, and so, without an explanation for their negative affect, they report *decreased* desire for control. It is possible that this is an attempt to compensate for the negative affect by distancing the self from causing the specific action or to reduce potential responsibility for the action. For participants in the nondisclosure explanatory vacuum, however, we may have evidence that there is a larger threat to personal control, as participants without an explanation for their negative affect report an increased desire for control (as predicted).

Another related, but slightly different possible explanation for the difference in desire for control in the two explanatory vacuum conditions is that the two may not be symmetrical, specifically in terms of acts of commission vs. acts of omission. In other words, in one condition participants commit an act that does not make sense in context (disclose to an untrustworthy experimenter) whereas those in the nondisclosure explanatory vacuum do *not* act, they withhold from a trustworthy experimenter. For those in the "commission" explanatory vacuum condition, an additional explanation for their discomfort (noise) prevents *decrease* in desire for control. For those in the "omission" explanatory vacuum condition, an additional explanation for their discomfort prevents an *increase* in desire for control. When one has already committed an action that one does not want responsibility for, it might make sense to decrease one's desire for control and responsibility over those actions; yet when one withholds action, perhaps it evokes a feeling of lack of control that must be compensated for or a feeling that there is still time to act.

Taken together, it appears that our two explanatory vacuum conditions have opposing effects on need for control. This raises interesting questions about the explanatory vacuum, and suggests that the type of behavior may matter, or even that unexplained negative (but not positive) behaviors increase need for control. Much research is needed in order to understand how this need relates to nonconsciously activated behavior. Our research so far attempted to uncover an interesting problem in the attribution of personal agency, namely, some actions have traceable roots partly outside of awareness and outside the boundaries of our heads; yet recognition of these actions is aversive.

Future Directions

We hope that by sharing these experiments, we will spark many questions relating to the perception of nonconsciously activated behavior, particularly as considered from the point of view of agency. First, future research would benefit from an experiment using simpler behaviors (e.g., arm movements; see Wegner, Sparrow, & Winerman, 2004) and a more direct, explicit threat to agency. In addition, further research would benefit from attempting to understand the function of confabulation in everyday life. Participants confabulate reasons for quotidian choices like board games and tutors when provoked (Bar-Anan, Wilson, & Hassin, 2010). While we had interpreted this as a general need to have a plausible reason for all behaviors, it now seems more likely that confabulation primarily serves to explain the origin of the specific behavior (and avoid the introspective blank). Future research would also benefit from attempting to understand what motives drive this type of confabulation behavior. For example, given the evidence of provoked confabulation, perhaps social norms or lay theories about actions (e.g., Wegner, 2002) dictate that having no reason for one's own behavior is socially unacceptable. One interesting question that this raises is whether or not, and how often, people confabulate in private. If confabulation primarily concerns social desirability, we would not expect any private confabulation of this kind at all. On the other hand, if people do confabulate reasons for their behavior privately, a number of possible motives may be at play.

One possibility is preference for consistency which has been previously associated with the explanatory vacuum (Parks-Stamm et al., 2010). Individuals who are high in preference for consistency may be more likely to notice or to care when their behavior does not match social norms or personal standards. For example, in one study (Parks-Stamm et al., 2010, Study 3), participants played a cooperative game with a partner and were given either a conscious or a nonconscious competitive goal. For those with a nonconscious goal, it was hypothesized that a lingering feeling of guilt would make participants want to act pro-socially toward the partner. To test this, a Dictator Game was played, in which they could decide how many lottery tickets to share with a partner. Preference for consistency

positively predicted the number of tickets shared in the Dictator Game only when norm-violating competitive behavior could not be explained by an earlier conscious goal. When the earlier conscious goal could explain the behavior, the opposite association was found between preference for consistency and tickets shared. This suggests that the more participants expect themselves to be consistent with their past goals, the less driven individuals are to engage in cooperative interpersonal behavior when an earlier goal can explain their antisocial behavior (i.e., when there's nothing to compensate for). The relationship between the explanatory vacuum and preference for consistency may suggest a strong link between the explanatory vacuum and cognitive dissonance. It would be interesting to examine what cognition is inconsistent with the behavior. For example, it is possible that participants observe their own behavior and find it contradicts self-standards (Stone & Cooper, 2001) or they infer a negative trait about themselves (Bem, 1967), which contradicts overall self-affirming motives (Steele, 1988). Of course, it is possible that confabulation can serve multiple motives in different contexts or that confabulation is not motivational in nature.

An important caveat to future research more generally is the distinction between desire for agency, predictability and violating expectations. Only agency specifically refers to the causal efficacy of the self (and from the first person perspective, the prediction of actions via intent). Violating expectations and unpredictability are not always negative. For instance, unexpected rewards are experienced as more pleasurable than expected ones (Wilson, Centerbar, Kermer, & Gilbert, 2005). Identifying an expectation violation may even be entertaining as in the case of watching someone perform magic tricks, and indeed, some have argued that humor functions as an internal reward system specifically tuned to catching expectation violations (Hurley, Dennett, & Adams, 2011). It would be interesting to find out whether some expectation violations that are self-generated are regarded as amusing as well.

Future research on the explanatory vacuum might also profit from insights of the research of Julius Kuhl and colleagues on PSI theory. First, Kuhl's PSI theory distinguishes between implicit and explicit affect, and Quirin, Kazen and Kuhl (2009) have recently developed an implicit positive and negative affect test (IPANAT). It seems possible that the negative affect we have observed as a central feature of the explanatory vacuum is more of an implicit than an explicit kind. After all, it is resulting from implicit self-regulation triggered by priming goals outside of conscious awareness. If this were true, assessing people's negative affective state while they are in an explanatory vacuum via the IPANAT might be the more appropriate way of testing the assumption that the explanatory vacuum is associated with negative affect. Finally, PSI theory offers a concept called self-infiltration (Baumann & Kuhl, 2003; see also Chapter 16 in this volume). It pertains to the phenomenon of falsely attributing externally controlled goals and activities to oneself. Research has shown that people in a bad mood who are high on state orientation are particularly vulnerable to this phenomenon. Applying the concept of self-infiltration to goal priming, it seems that a state-oriented individual when in a bad mood may be comparatively more susceptible to goal priming. And even more interesting, would they be more or less likely to confabulate plausible explanations for the behaviors they performed in the service of goals activated outside of their awareness. In summary, the explanatory vacuum is an intriguing phenomenon that investigates the interaction between two self-regulatory states (consciousness and nonconsciously activated behaviors); future research on the explanatory vacuum would greatly benefit from the integration of Kuhl's theories and measurements.

Corresponding author
Ana P. Gantman
agantman@nyu.edu

References

Adriaanse, M. A., Kroese, F. M., Weijers, J., Gollwitzer, P. M., & Oettingen, G. (2016). *Explaining unexplainable food choices*. Manuscript submitted for publication.

Adriaanse, M. A., Weijers, J., De Ridder, D. T., 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. http://doi.org/10.1002/ejsp.2005

Bar-Anan, Y., Wilson, T. D., & Hassin, R. R. (2010). Inaccurate self-knowledge formation as a result of automatic behavior. *Journal of Experimental Social Psychology, 46*, 884–894. http://doi.org/10.1016/j.jesp.2010.07.007

Bargh, J. A. (1990). Auto-Motives: Preconscious determinants of social interactions. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition* (Vol. 2, pp. 93–130). New York, NY: Guilford Press.

Bargh, J. A., & Gollwitzer, P. M. (1994). Environmental control of goal-directed action: Automatic and strategic contingencies between situations and behavior. In W. D. Spaulding (Ed.), *Integrative views of motivation, cognition, and emotion* (pp. 71–124). Lincoln, NE: University of Nebraska Press.

Bargh, J. A., Gollwitzer, P. M., Lee-Chai, A., Barndollar, K., & Trötschel, R. (2001). The automated will: nonconscious activation and pursuit of behavioral goals. *Journal of Personality and Social Psychology, 81*, 1014–1027. http://doi.org/10.1037/0022-3514.81.6.1014

Bargh, J. A., Gollwitzer, P. M., & Oettingen, G. (2010). Motivation. In S. Fiske, D. Gilbert, & G. Lindzey (Eds.), *Handbook of social psychology* (5th ed., pp. 268–316). New York, NY: Wiley.

Bargh, J. A., Schwader, K. L., Hailey, S. E., Dyer, R. L., & Boothby, E. J. (2012). Automaticity in social-cognitive processes. *Trends in Cognitive Sciences, 16*, 593–605. http://doi.org/10.1016/j.tics.2012.10.002

Baumann, N., & Kuhl, J. (2003). Self-infiltration: Confusing assigned tasks as self-selected in memory. *Personality and Social Psychology Bulletin, 29*, 487–497. http://doi.org/10.1177/0146167202250916

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review, 74*, 183–200. http://doi.org/10.1037/h0024835

Berlyne, N. (1972). Confabulation. *The British Journal of Psychiatry, 120*, 31–39. http://doi.org/10.1192/bjp.120.554.31

Brehm, J. (1956). Post-decision changes in desirability of alternatives. *Journal of Abnormal and Social Psychology, 52*, 384–389. http://doi.org/10.1037/h0041006

Burger, J. M., & Cooper, H. M. (1979). The desirability of control. *Motivation and Emotion, 3*, 381–393. http://doi.org/10.1007/BF00994052

Chartrand, T. L. (2005). The role of conscious awareness in consumer behavior. *Journal of Consumer Psychology, 15*, 203–210. http://doi.org/10.1207/s15327663jcp1503_4

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

Cooney, J. W., & Gazzaniga, M. S. (2003). Neurological disorders and the structure of human consciousness. *Trends in Cognitive Sciences, 7*, 161–165. http://doi.org/10.1016/S1364-6613(03)00058-5

Dennett, D. C. (1993). *Consciousness explained.* London, UK: Penguin.

Dijksterhuis, A., & Aarts, H. (2010). Goals, attention, and (un) consciousness. *Annual Review of Psychology, 61*, 467–490. http://doi.org/10.1146/annurev.psych.093008.100445

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. http://doi.org/10.1037/0022-3514.67.3.382

Festinger, L. (1962). *A theory of cognitive dissonance* (Vol. 2). Paolo Alto, CA: Stanford University Press.

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. http://doi.org/10.1037/0022-3514.84.2.296

Fitzsimons, G. M., & Bargh, J. A. (2003). Thinking of you: Nonconscious pursuit of interpersonal goals associated with relationship partners. *Journal of Personality and Social Psychology, 84*, 148–164. http://doi.org/10.1037/0022-3514.84.1.148

Gollwitzer, P. M., Sheeran, P., Trötschel, R., & Webb, T. (2011). Self-regulation of behavioral priming effects. *Psychological Science, 22*, 901–907. http://doi.org/10.1177/0956797611411586

Harmon-Jones, E., Harmon-Jones, C., Amodio, D. M., & Gable, P. A. (2011). Attitudes toward emotions. *Journal of Personality and Social Psychology, 101*, 1332–1350. http://doi.org/10.1037/a0024951

Hassin, R. R., Uleman, J. S., & Bargh, J. A. (Eds.). (2004). *The new unconscious*. Oxford, UK: Oxford University Press.

Hirstein, W. (2005). *Brain fiction: Self-deception and the riddle of confabulation*. Cambridge, MA: MIT Press.
Hurley, M. M., Dennett, D. C., & Adams, R. B. (2011). *Inside jokes: Using humor to reverse engineer the mind*. Cambridge, MA: MIT press.
Koole, S. L., McCullough, M. E., Kuhl, J., & Roelofsma, P. H. (2010). Why religion's burdens are light: From religiosity to implicit self-regulation. *Personality and Social Psychology Review, 14*, 95–107. http://doi.org/10.1177/1088868309351109
Kopelman, M. D. (1987). Two types of confabulation. *Journal of Neurology, Neurosurgery & Psychiatry, 50*, 1482–1487. http://doi.org/10.1136/jnnp.50.11.1482
Kuhl, J. (2000). A functional-design approach to motivation and self-regulation: The dynamics of personality systems and interactions. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds), *Handbook of self-regulation* (pp. 111–169). San Diego, CA: Academic Press.
Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science, 12*, 135–140. http://doi.org/10.1111/1467-9280.00323
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*, 231–259. http://doi.org/10.1037/0033-295X.84.3.231
Oettingen, G., Grant, H., Smith, P. K., Skinner, M., & Gollwitzer, P. M. (2006). Nonconscious goal pursuit: Acting in an explanatory vacuum. *Journal of Experimental Social Psychology, 42*, 668–675. http://doi.org/10.1016/j.jesp.2005.10.003
Parks-Stamm, E. J., Oettingen, G., & Gollwitzer, P. M. (2010). Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving. *Journal of Experimental Social Psychology, 46*, 531–542. http://doi.org/10.1016/j.jesp.2010.02.004
Proulx, T., Inzlicht, M., & Harmon-Jones, E. (2012). Understanding all inconsistency compensation as a palliative response to violated expectations. *Trends in Cognitive Sciences, 16*, 285–291. http://doi.org/10.1016/j.tics.2012.04.002
Quirin, M., Kazén, 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. http://doi.org/10.1037/a0016063
Shah, J. Y., & Kruglanski, A. W. (2003). When opportunity knocks: bottom-up priming of goals by means and its effects on self-regulation. *Journal of Personality and Social Psychology, 84*, 1109–1122. http://doi.org/10.1037/0022-3514.84.6.1109
Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology, 21*, 261–302.
Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology, 37*, 228–243. http://doi.org/10.1006/jesp.2000.1446
Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. *Proceedings of the National Academy of Sciences, 109*, 8038–8043. http://doi.org/10.1073/pnas.1202129109
Tanner, R. J., Ferraro, R., Chartrand, T. L., Bettman, J. R., & Van Baaren, R. (2008). Of chameleons and consumption: The impact of mimicry on choice and preferences. *Journal of Consumer Research, 34*, 754–766. http://doi.org/10.1086/522322
Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge, MA: MIT Press.
Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: experiencing control over the movements of others. *Journal of Personality and Social Psychology, 86*, 838– 848. http://doi.org/10.1037/0022-3514.86.6.838
Wheatley, T. (2009). Everyday confabulation. In W. Hirstein (Ed.), *Confabulation: Views from neuroscience, psychiatry, psychology, and philosophy* (pp. 205–225). New York, NY: Oxford University Press.
Wilson, T. D., Centerbar, D. B., Kermer, D. A., & Gilbert, D. T. (2005). The pleasures of uncertainty: prolonging positive moods in ways people do not anticipate. *Journal of Personality and Social Psychology, 88*, 5–21. http://doi.org/10.1037/0022-3514.88.1.5
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. http://doi.org/10.1037/h0036651