17

Justification as antecedent and consequence of self-control failure

Marieke Adriaanse and Sosja Prinsen

In this chapter, we will introduce the topic of justification and discuss how justification processes may precede incidents of self-control failure – when our actions are inconsistent with our long-term (global) goals and values (Fujita, 2011) – or follow behaviors that signal self-control failure. Specifically, in the first half of this chapter, we will present research showing that justification processes can be a cause of self-control failure when people engage in *self-licensing*. That is, we will discuss evidence from the moral, consumer, and health domain showing that people may deliberately engage in behaviors that violate long-term goals when the context justifies doing so. At the end of this section, we will discuss the long-term effects of self-licensing and call for future research to explore whether self-licensing is ultimately harmful or adaptive. In the second half of this chapter, we will present evidence demonstrating that justification processes may also be a consequence of self-control failure. Specifically, we will argue that people have a tendency to justify their behavior by *confabulating* reasons for their behavior when it is perceived as self-control failure and the actual cause of the behavior is inaccessible. We will briefly discuss the background of the term confabulation ("to lie without the intent to deceive," Hirstein, 2009) and discuss recent studies demonstrating how confabulation may be of relevance in the domain of social and health psychology. In addition, we will theorize how confabulation may affect future self-control attempts.

Justification as a cause of self-control failure: self-licensing

In the literature on self-control, self-control failure is often explained as being the result of impulses taking precedence over reflective considerations (e.g., Hagger, Wood, Stiff, & Chatzisarantis, 2010; Hofmann, Friese, & Wiers, 2008). From this perspective, being confronted with a temptation activates our impulses to fulfill this momentary desire, which can only be tempered if our capacity to reason and reflect on our actions functions properly. For example, after a demanding workday, people are left with less cognitive capacity to steer oneself to the gym and away from the TV, making the realization of the intention to exercise less likely. Yet, in the research on self-licensing which we will review below, the assumption that our reasoning will guide us in the direction of behavior that aligns with long-term goals is challenged. These studies have demonstrated that our reasoning can just as well be motivated by our desires and lead us into temptation. That is, research on self-licensing shows that long-term goals can be deliberately violated when

the context justifies doing so (e.g., De Witt Huberts, Evers, & De Ridder, 2014). For example, instead of hampering our ability to act responsibly, a long workday can also be intentionally used as a reason to justify skipping the gym and watching TV instead ('I deserve it'). Actually, people may even actively seek for reasons that justify such violations, especially when confronted with readily available temptations. Thereby, self-licensing provides a new and complementary perspective on self-control failure by stating that succumbing to temptation is not necessarily always the result of uncontrollable impulses, but it can also stem from reasoning processes that are typically associated with successful self-control.

Self-licensing was first introduced in the domain of moral behavior (Monin & Miller, 2001), proposing the notion that people feel more free to act immorally after an initial moral act, like being more likely to cheat and steal after purchasing green (vs. conventional) products (Mazar & Zhong, 2010). However, people do not always have to act morally for this effect to occur; merely recalling or imagining a moral act has been found to result in moral licensing effects as well. For example, people were found to be less willing to donate money after recalling a situation of helping other people (Jordan, Mullen, & Murnighan, 2011) or imagining supporting a foreign student (Khan & Dhar, 2006). Overall, a meta-analysis of 91 moral licensing studies established an estimated effect size of $d = .31$, suggesting a small-to-medium moral licensing effect (Blanken, Van de Ven, & Zeelenberg, 2015). Altogether, these findings support the proposition that people who behave in a good (moral) way later feel that they are permitted to engage in undesirable (immoral) behavior (Monin & Miller, 2001).

After its introduction, self-licensing was studied in other domains as well, like consumer and health behavior. With this came a more general conceptualization of self-licensing, which is "the act of making excuses for one's discrepant behavior before actual enactment, such that the prospective failure is made acceptable for oneself" (De Witt Huberts et al., 2014, p. 121). An important difference with moral licensing is that this definition states that self-licensing is triggered by a self-control dilemma; i.e., a decision between an immediately gratifying option (e.g., spending money on luxury goods) and an option with direct costs but long-term benefits (e.g., saving money for retirement). This dilemma triggers the need to find an excuse or reason that justifies going for the indulgent rather than prudent choice. Thus, self-licensing is not only about being more likely to give in to temptation in response to feelings of deservingness after having behaved responsibly, but also encompasses active engagement in using and searching for available justifications. This is nicely illustrated in a study by Effron, Monin, and Miller (2013). In this study, participants were asked to provide a consumer product evaluation (cover-story), consisting of several tasting and rating sessions of different types of snacks. First, they were asked to choose a set of unhealthy snacks to taste in the second tasting session. After choosing one of two sets, participants in the experimental condition were steered toward choosing an unhealthy snack for the first tasting session: participants could choose between either eating two cloves of raw garlic, or four freshly baked cookies. As can be expected, 91% of participants chose the latter option. Importantly, in a pilot test it was found that participants would anticipate feeling guilty when doing so. In the control condition, participants were told they only needed to examine, but not eat, both the garlic and cookies. Then, all participants rated the healthiness of both sets of snacks that they chose earlier. The unchosen snacks were rated as unhealthier than the chosen snacks in the experimental condition, but received equal ratings in the control condition. For the chosen snacks, no difference was observed between conditions. This suggests that participants who were tempted to choose an unhealthy snack 'strategically evaluated' the snacks that they declined to eat just before as more unhealthy, compared to participants who were not tempted. Thus, only when facing temptation, participants showed a need to exaggerate the unhealthiness of forgone snacks, as a means to justify a subsequent indulgent choice.

Eating behavior, specifically choosing what and how much to eat, is a typical example of behavior that involves dealing with self-control dilemmas. Hence, a growing number of studies examine food choices and food intake to establish self-licensing effects. Empirical studies have shown, for example, that participants were more likely to choose an unhealthy (e.g., chocolate bar) over a healthy snack (e.g., apple) when they recalled an altruistic action (Weibel, Messner, & Brügger, 2014); recalled a personal achievement (Wilcox, Kramer, & Sen, 2011); or believed that they had made sufficient progress toward their weight loss goal (Fishbach & Dhar, 2005). In this latter study, the perceived discrepancy between participants' current weight and goal weight was manipulated to look either small or large. Participants were asked to report their current weight in a textbox in the center of a scale that had either −5 kg and +5 kg or −20 kg and +20 kg as its endpoints. Then they indicated their goal weight by coloring the arrow that extended outward to the left (to endpoint −5 or −20 kg) to the point that represented their goal weight. On the narrow (−5 kg) scale, a goal weight of 3 kg less than one's current weight would mean coloring 60% of the scale. In contrast, on the wide (−20 kg) scale, wanting to lose 3 kg would result in coloring only 15% of the scale. Hence, the visual discrepancy between one's current and goal weight would appear smaller on the wide scale compared to the narrow scale. It was expected that a small perceived discrepancy would induce a sense of being closer to one's goal weight compared to a large perceived discrepancy. The results showed that 85% of participants in the wide-scale condition subsequently chose a chocolate bar over an apple, compared to 58% of participants in the narrow-scale condition. This suggests that perceiving the goal discrepancy as small licensed participants to choose an unhealthy rather than healthy snack, as it signaled sufficient progress toward the goal of losing weight.

In addition to unhealthy food choices, increases in unhealthy food intake have also been observed to result from justification processes. In a study by De Witt Huberts, Evers, and De Ridder (2012), participants who were led to believe that they invested more effort on a task (effort condition) subsequently ate more unhealthy snacks in a 'taste test' than participants in the control condition. In this study, all participants performed the same task of typing the first letter of each word that appeared on a computer screen. Participants in the control condition did this for 10 consecutive minutes, whereas participants in the effort condition were told after the first 5 minutes that the task was finished, but that some participants would be randomly selected to do the task again (in reality, all participants completed the exact same task). So although actual effort was kept constant, perceived effort differed between conditions.

In another study by Taylor, Webb, and Sheeran (2013), female students were primed to justify indulgence. They were told to imagine themselves in a situation where they decide to go on holiday with friends rather than with their boyfriend, and to write down as many reasons as they could think of to justify this decision (e.g., 'I'll make it up to my boyfriend'). In the control condition, participants were asked to write down possible destinations for a holiday with friends. Afterwards, it was found that participants who previously generated reasons ate more of an unhealthy snack than participants who did not receive this justification prime. Importantly, this effect was only observed for participants who reported strong intentions to decrease their unhealthy snack intake. This finding suggests that, paradoxically, individuals with the strongest intentions are also the ones who are most susceptible to justify indulgence.

These studies aptly illustrate that there is a wide variety of justifications, but also that self-licensing effects are not domain specific. That is, behaving morally, like being altruistic, does not only license subsequent immoral behavior (Jordan et al., 2011), but unhealthy food choices as well (Weibel et al., 2014). Studies in the domain of consumer behavior have found similar cross-domain effects, by demonstrating that participants were more likely to choose luxury (e.g., expensive perfume) over necessity goods (e.g., vacuum cleaner) when they had just committed to

a charitable act (e.g., Khan & Dhar, 2006). Accordingly, it has been proposed that "when people find themselves in a situation where they are tempted by something they know they really should not do, they might be successful in constraining themselves, unless they find a reason, any reason, to give in" (De Witt Huberts et al., 2014, p. 122). Hence, as long as the reason seems valid to the person using it, it can be used to justify goal-violating behavior.

Several mechanisms have been proposed to underlie self-licensing effects. These include reinforced self-concept, motivated reasoning, prefactual cognitive dissonance, and anticipated affect (see De Witt Huberts et al., 2014, for an overview). One mechanism that logically follows from most of the previously discussed studies is reinforced self-concept. By performing 'good' behavior, like acting altruistically or not eating unhealthy snacks, a positive self-concept is reinforced. Subsequently, ensuing 'bad' behavior becomes less threatening as there is good behavior to buffer its negative impact on one's self-perceptions. So, eating a cookie becomes more acceptable after previously resisting a cookie, as the latter has proven that you are able to control yourself. Motivated reasoning is also driven by the need to preserve one's self-concept, but instead of wanting to preserve a favorable view of oneself, motivated reasoning is about perceiving oneself as a rational person. By employing reasons to justify discrepant behavior, the illusion of acting rationally is maintained, even though the behavior can be considered irrational from an objective perspective. For example, spending your savings on a pair of new shoes seems more rational after having first convinced yourself that all your other shoes are totally out of style.

Closely related to motivated reasoning is prefactual cognitive dissonance, in which people feel conflicted about the prospective goal violation and try to resolve this conflict by rationalizing their behavior *before* performing it ('pre facto'). This mirrors classic cognitive dissonance theory (Festinger, 1957) and research on confabulation (e.g., Adriaanse et al., 2014, discussed in the second half of this chapter), which both illustrate that the discomfort experienced *after* performing discrepant behavior prompts the need to rationalize the behavior. Hence, upon foreseeing the discomfort that would follow a certain indulgent behavior, self-licensing processes are triggered. This foreseeing is also central to anticipated affect, but rather than anticipating a general sense of discomfort, this account of self-licensing focuses on anticipated negative affective states like regret and guilt.

Similar to prefactual cognitive dissonance, expecting to feel guilty motivates people to prevent this from happening. Justifications can be an effective way to do so. This mechanism resembles the entitlement route to justification that has been put forward by Kivetz and Zheng (2006), which is based on the notion that people do not only suffer from a lack of self-control, but also from 'overcontrol' that prevents them from enjoying indulgent behaviors when they lack a sense of entitlement or deservingness for doings so. Hence, self-gratification is said to inherently evoke guilt, unless it is deserved through 'hard work' or some kind of achievement.

Evidently, these mechanisms are not mutually exclusive as they share similarities and overlap to some extent. Importantly, all mechanisms serve to resolve of a self-control dilemma in favor of the tempting option in a satisfactory manner. An important next step is to provide empirical evidence for these potential explanations. Whereas anticipated affect has been studied in the context of the entitlement route to justification (e.g., Kivetz & Zheng, 2006), and reinforced self-concept has been studied as a potential mechanism underlying self-licensing effects to some extent (e.g., Khan & Dhar, 2006), other explanations are yet to be directly examined.

Long-term effects of self-licensing

In addition to a lack of sufficient insight into the underlying mechanism, self-licensing research so far has mainly focused on immediate and single outcomes. That is, indulgent behavior is

usually assessed directly after participants have been experimentally manipulated to self-license, and often only once. Therefore, little is known about how self-licensing influences subsequent decision making, i.e., the sequence of choices that we make over longer periods of time. People often encounter multiple self-control dilemmas. It has, for example, been estimated that we already make 200 food-related decisions per day (Wansink & Sobal, 2007). This most likely applies to other behaviors as well: just imagine making a shopping trip to the mall or dealing with the constant lure of Facebook throughout the workday. Importantly, this emphasizes the need for looking beyond single outcomes and looking at behavioral patterns instead. After all, it is crucial to realize that one indulgent choice does not seriously harm the attainment of a long-term goal, whereas repeated indulgent choices over time do. Spending 5 minutes on Facebook does not necessarily interfere with your work performance, but it likely does when all these short moments add up to a few hours. Hence, a lack of insight into how self-licensing affects repeated decision making renders the conclusion that self-licensing leads to self-control failure, in terms of failing to achieve a long-term goal, premature – particularly as there may be a positive side to self-licensing in the long run that is currently overlooked.

When focusing on long-term outcomes, it may be observed that allowing oneself the occasional goal violation – by means of self-licensing – may be a better strategy than aiming for complete control over one's behavior. Considering the vast literature on self-control failure, it seems reasonable to suggest that it is impossible for people to always control their behavior. Also, people usually have multiple goals which inevitably requires some degree of failure in pursuing one goal to preserve the attainment of another. Hence, also in maintaining a proper balance between goals (see Fishbach & Dhar, 2005), goal violations can occur. So, if people then do violate a goal, it may be better to perceive this behavior as justified rather than experiencing it as failure. Indeed, it has been found that the latter negatively affects the handling of subsequent self-control challenges (Zemack-Rugar, Corus, & Brinberg, 2012). Moreover, there is initial evidence suggesting that this has to do with how goal violations are incorporated into one's self-perceptions. In a vignette study, it was demonstrated that when participants imagined themselves in a situation where they violated their diet with a license, they subsequently reported higher feelings of self-efficacy than participants who did not have a license for this transgression (Prinsen, Evers, & De Ridder, 2016). Importantly, this may, in turn, influence how subsequent temptations are handled. Thus, for future research it is important to address how self-licensing effects unfold over time and the role of self-perceptions regarding one's ability to deal with temptations in this.

Besides the theoretical relevance of looking at long-term outcomes, this knowledge is also pivotal to the development of appropriate intervention methods that aim to target self-licensing processes. As there is currently not enough insight into the role of justification in successful goal striving, there is little direction for how to approach this issue. A first step could be to identify under what conditions self-licensing can be considered threatening to goal striving (e.g., when individuals justify goal-violating behavior too often or too easily) and in what way self-licensing can be conducive to successful self-control (e.g., when self-perceptions of one's self-control capacity are protected despite an occasional goal violation). Identifying these conditions can be helpful to very precisely target behavior components that require either reinforcement or modification. Importantly, interventions directed at self-licensing processes can complement current behavior change techniques that mainly focus on impulsive determinants of self-control failure.

Summary

To conclude, goal violations are often explained in terms of not being able to resist temptations. However, people may deliberately choose to temporarily abandon their goal by employing

licenses that justify this discrepant behavior. This is referred to as self-licensing, and is driven by self-control dilemmas that necessitate choosing between two opposing options. At first sight it seems that self-licensing harms successful goal pursuit, as it makes it easier to choose the immediate gratifying yet irresponsible option. However, to establish whether such justification processes are ultimately harmful or beneficial, the long-term outcomes need to be examined.

Justification as a consequence of self-control failure: confabulation

In the section above, we argued that self-control failure can be caused by self-licensing, or justification processes, whereby individuals 'allow' themselves to deliberately act against their goals. In the section below, we will discuss how justification processes may also be of relevance in the *aftermath* of self-control failure. In doing so, we will return to the 'classic' case of self-control failure as an unwanted, unintentional behavior that occurs because we simply cannot exert self-control all of the time. That is, regardless of whether one perceives self-control as a limited resource that gets depleted (e.g., Baumeister, Bratslavsky, Muraven, & Tice, 1998) or adopts a motivational account of self-control (see Milyavskaya & Inzlicht, Chapter 2 of this volume; Molden et al., Chapter 11 of this volume), it is a consistent finding that people's exertion of self-control and their success in controlling their impulses fade over time (see Milyavskaya & Inzlicht, Chapter 2 of this volume). Similarly, it is a well-documented finding that our attempts to act in line with our goals are frequently outrun by everyday nonconscious or impulsive influences on behavior (Hofmann, Friese, & Wiers, 2008). Below, we will argue that in many of these cases where we are confronted with such inevitable instances of self-control failure, we experience a need to justify this behavior (Oettingen, Grant, Smith, Skinner, & Gollwitzer, 2006). Although there are certainly instances of self-control failure that can be straightforwardly explained because the reasons for engaging in the 'wrong' behavior are clear (e.g., when a person is forced to choose between two competing goals or when self-control failure is the consequence of self-licensing), often the reasons are ambiguous or even inaccessible (Nisbett & Wilson, 1977; Sheeran, Gollwitzer, & Bargh, 2013).

Consider the example of Anna, a dieter who finds herself ordering a big piece of chocolate cake when meeting her friend at a local restaurant. It is quite likely that this instance of self-control failure is triggered by cues – the influence of which Anna is unaware, such as food advertisements in the restaurant, or her friend's eating behavior. Indeed, in contrast to the lay belief that our actions are a consequence of our conscious will (Aarts, Custers, & Wegner, 2005; Renes & Aarts, Chapter 16 of this volume; Wegner & Wheatley, 1999), much of our daily behavior appears to be the consequence of processes occurring outside of conscious awareness (Sheeran et al., 2013). For example, it has been shown that TV food advertisements can automatically increase snack intake (Harris, Bargh, & Brownell, 2009) and social cues, such as other people's eating behavior, can influence food intake through mimicking processes of which people are completely unaware (Tanner, Ferraro, Chartrand, Bettman, & van Baaren, 2008).

Regardless of whether people are aware or unaware of the presence of these cues, and even when they accept the idea that these cues do have an influence on their behavior, people frequently have limited introspective awareness of the degree of influence exerted by these cues in specific instances (Chartrand, 2005; Molden, 2014; Nisbett & Wilson, 1977). Interestingly, in these cases where people have limited insight into the causes of their behavior (i.e., the behavior is the result of processes of which the actor is – at least partly – unaware) and the behavior demands an explanation (i.e., it is inconsistent with the actor's long-term goals), people do not admit that they do not know (Nisbett & Wilson, 1977), or are uncertain about why they behaved the way they did. Rather, people appear to have a strong tendency to *confabulate* explanations

for their behavior when the actual cause is inaccessible (e.g., Adriaanse, Weijers, De Ridder, De Witt Huberts, & Evers, 2014; Bar-Anan, Wilson, & Hassin, 2010; Parks-Stamm, Oettingen, & Gollwitzer, 2010). For example, Anna may justify eating the chocolate cake by confabulating that the reason for her indulgent behavior was that she has had a very stressful day. Note that our hypotheses do not require that people are completely unaware of the cause of their behavior as long as they underestimate the extent to which a certain process or cue affected their behavior. In addition, similar processes are expected to occur if people simply forget the cause of their actions.

In this chapter, confabulation is defined as the process of adopting an erroneous reason for one's behavior "without the intent to deceive and without knowing that this claim is ill-grounded" (Hirstein, 2009). The term confabulation was first used in the early 1900s by neurologists to describe false claims presented as memories by patients with Korsakoff syndrome (Hirstein, 2009). Gradually, the definition of confabulation was used more broadly to cover false claims by patients with other disorders, such as split-brain patients or patients with misidentification syndrome. Nowadays, the term confabulation is used also in the study of action and intention in healthy samples, where confabulations serve the purpose of restoring a sense of agentic coherence and consistency (Wheatley, 2009). In this chapter, we will review a specific case of this type of confabulation; that is, confabulation that occurs upon being confronted with behavior that is triggered by processes occurring outside of conscious awareness.

Crucially, not all behaviors of which the causes are unknown are expected to trigger confabulation. As mentioned above, confabulation is expected to occur when the cause of the behavior is inaccessible *and the behavior demands an explanation*. So, when does a behavior 'demand an explanation'? In the clinical literature on confabulation, a distinction is made between confabulations formed reflexively, labeled 'spontaneous confabulations', and so-called 'provoked confabulations', which are given in response to a question by an authority figure (Kopelman, 1987). In the latter situation, the demand for an explanation stems from the explicit request to explain one's actions by the experimenter or an authority figure (see also Bar-Anan, Wilson, & Hassin, 2010; Nisbett & Wilson, 1977). In the present chapter, we will focus on s*pontaneous confabulation*, and argue that confabulation is a relevant psychological process which arises spontaneously without interference of an experimenter (or any other person) in daily life. In these cases, there is no direct request for an explanation, but the demand for an explanation stems from the experience of inconsistency between one's standards (a term we use in the wider sense to denote global goals, values, norms, attitudes etc.) and the observed behavior (Oettingen et al., 2006; Parks-Stamm et al., 2010). In other words, these are situations where the behavior that is enacted signals self-control failure.

Whenever our standards and behavior do not align, our need to maintain consistency is jeopardized, which is experienced as aversive, as is well documented in the literature on cognitive dissonance (e.g., Elliot & Devine, 1994; Cooper & Fazio, 1984; Festinger, 1957; Stone & Cooper, 2001). When experiencing such an inconsistency, people are motivated to confabulate an explanation for their behavior to reduce the negative affect associated with this perceived inconsistency (Adriaanse et al., 2014). Coming back to our example of Anna, looking down on the crumbs on her plate, she is confronted with an inconsistency between her behavior (eating a calorie-rich cake) and her dieting goal. As Anna has no insight into the cause of her behavior, she is left with an unexplainable inconsistency that is experienced as unpleasant. This negative feeling in turn motivates her to search for explanations and leaves her concluding that her indulgent behavior must be the result of a very stressful day at work.

Oettingen and colleagues (2006) provided empirical evidence for the notion that being confronted with unexplainable behavior that violates a personal standard (i.e., 'acting in an explanatory vacuum', Oettingen et al., 2006) is experienced as unpleasant. These authors showed

that not having an explanation for one's behavior because this behavior is the result of non-conscious processes triggered negative affect, but only when the behavior violated a salient norm and therefore *demanded* an explanation. Specifically, in a series of studies, they showed that participants who were subliminally primed to violate a social norm, but not participants who violated this norm as a result of a consciously provided goal, experienced elevated levels of negative affect. Crucially, negative affect did not increase when participants were primed (or consciously instructed) to act in line with the social norm.

In a second line of studies, Parks-Stamm and colleagues (2010) provided initial evidence for the notion that the negative affect that results from acting in an explanatory vacuum is not the endpoint, but rather triggers a tendency to confabulate. Building on the notion that negative affect is an aversive state that people are motivated to reduce (Stone & Cooper, 2001), Parks-Stamm and colleagues (2010) tested whether providing people with a plausible explanation for their norm-violating behavior would reduce the level of experienced negative affect. This plausible explanation was made accessible to participants at the onset of the experiment by asking them to pursue a goal that could later on be used as a potential source of misattribution. Indeed, results indicated that participants who were primed to act in a way that violated a salient norm felt significantly less negative when this behavior was in line with a conscious goal provided to participants earlier in the experimental procedure. This finding implies that people are likely to 'use' an available plausible explanation to reduce the negative affect associated with the lack of such an explanation.

More direct evidence for the relevance of this process to the domain of self-control comes from recent work by Adriaanse and colleagues (2014). In a first study, participants were manipulated to behave more or less prosocially by exposing them to neutral or aggressive video games. Next, prosocial behavior was assessed: Participants were asked to help out a fellow student (which was assumed to be the norm for the participating students) by completing as many trials as possible on a tedious computer task – which can been considered an act of self-control – and to stop when they felt they had sufficiently helped. In line with the manipulation, participants who had played the aggressive game completed fewer help trials than participants who had played a neutral game. As expected, the participants in the aggressive game condition subsequently experienced an increase in negative affect, which in turn predicted a more negative evaluation of the lab space in which they performed the task (e.g., 'the chair was uncomfortable'). These lower, more negative evaluations of the lab were interpreted as confabulated reasons for quitting sooner on the voluntary task. In other words, it was concluded that participants justified their lower levels of helping behavior by rating the lab as less pleasant or comfortable.

Taking these findings to the health domain, these authors then conducted a second study (Adriaanse et al., 2014, Study 2) in which a lexical decision task was used to prime a neutral (control condition) or a hedonic goal (hedonic priming condition). Participants were then asked to eat as much chocolate as they wanted in a subsequent so-called 'taste test'. To manipulate whether indulging in chocolates was experienced as violating personal standards, the study included people who at baseline had indicated having either high or low dieting standards. After the taste test, in which the primed participants consumed more chocolate than the other participants, all participants were exposed to a text suggesting that cognitively demanding tasks increase cravings for sugar. Thus, participants were exposed to a potential excuse for their apparent self-control failure by explaining to them that it is normal to consume sugar (glucose)-rich foods after having performed a cognitively exhausting task (such as the lexical decision task). When participants were subsequently asked to provide feedback on the lexical decision task that had preceded the chocolate tasting, participants in the hedonic priming condition (who had indeed consumed more chocolate) rated this task as more cognitively exhausting, but only when

they had high dieting standards. So, participants who, despite trying to restrict their unhealthy food intake, indulged in chocolate without having a good explanation for doing so subsequently attributed this apparent act of self-control failure to having been cognitively exhausted by performing the lexical decision task prior to eating. The effect on confabulation was mediated by higher scores on negative affect after the taste test for participants with high dieting standards in the hedonic priming condition. Together, the results were indicative of mediated moderation; both the direct and indirect (via negative affect) effects of performing an unconsciously activated behavior on confabulation were moderated by personal standards.

Recently, this study was replicated with the addition of a 'prime and tell' condition to rule out the alternative explanation that the different evaluations of the lab or the lexical decision task reflect a mere mood congruent memory bias (as people scoring lower on these evaluations were also experiencing negative affect) rather than an attempt to actually explain or justify the observed behavior (Adriaanse, Kroese, Weijers, Gollwitzer, & Oettingen, in press). The crucial difference between the prime and tell condition and the regular hedonic priming condition was that participants in the first condition were provided with an explanation for their behavior after completing the taste test. Participants in this condition, for whom there was no longer a reason to explain the behavior, did not display the same tendency to report being more cognitively exhausted after performing the lexical decision task. This was taken as evidence that the ratings of cognitive exhaustion after performing the lexical decision task for participants with high dieting standards in the hedonic prime condition were indeed indicative of confabulation. In summary, the aforementioned studies suggest that behavior for which participants have no accessible explanation and that is perceived of as signaling self-control failure as it is inconsistent with personal standards (e.g., behavior that violates social norms or dieting standards) leads to increased feelings of negative affect, and a subsequent tendency to confabulate a reason for this behavior.

These findings also suggest that people are quick to take up any explanation subtly suggested to them when they experience unexplainable instances of self-control failure. Although in these studies the opportunity to confabulate was provided to them (e.g., by providing a text or asking for an evaluation), still, participants were not probed to justify or explain their behavior whatsoever, suggesting that confabulation arose relatively spontaneously. Indeed, other studies demonstrated that confabulation arises regardless of whether participants are provided with the opportunity to consciously reflect on the behavior that was executed, indicating that confabulation occurs reflexively rather than reflectively (Parks-Stamm, Oettingen, & Gollwitzer, 2010). Future research, using more sophisticated measures, is necessary to provide more convincing evidence for the notion that confabulation arises truly reflexively and spontaneously and without any interference of an experimenter whatsoever.

It should be noted that, even if confabulation is a relatively automatic process, this still does not mean that any reason will be automatically accepted as an explanation for apparent self-control failure. It is to be expected that confabulations are more likely to depend on, for example, the plausibility of the reasons that are accessible to use (Nisbett & Wilson, 1977; Tversky & Kahneman, 1974). This is exactly what was suggested in a recent study by Adriaanse, Prinsen, De Witt Huberts, De Ridder, and Evers (2016; note that these results should be considered preliminary as the limited power warrants replication studies with larger samples). On the first day of this two-day study, participants who had either high or low self-reported emotional eating tendencies watched a neutral video and reported their level of negative affect afterwards. Participants were then provided with four types of snacks and were instructed to eat what they thought was 20 g of each snack. On the second day, participants were randomly assigned to one of two bogus feedback conditions in which they were told that they either ate roughly

the prescribed amount of each snack or that they had eaten way more than prescribed. After receiving this false feedback, participants retrospectively reported on their affective state after watching the video, which they had viewed just before the snack estimation task. Despite no differences in the negative affect reported right after watching the video, participants who had reported that they perceived themselves as emotional eaters, and who were told that they had eaten more than the norm, retrospectively (i.e., one day after watching the video) described themselves as feeling more negatively after watching the video, before eating. In other words, only those participants for whom emotions represented a highly plausible reason for overeating were inclined to post-hoc attribute overeating to the experience of negative emotions prior to participating in the estimation task.

The above theorizing and experiments are strongly related to work on cognitive dissonance (Festinger, 1957; Stone & Cooper, 2001; Harmon-Jones, Harmon-Jones, & Levy, 2015). Yet, while the general process of experiencing discomfort related to being confronted with inconsistency is similar, cognitive dissonance and confabulation also differ by proposing two distinct routes to account for behaviors that are inconsistent with personal standards (which, in our broad definition, includes the concept of attitudes). That is, despite the fact that attitude change was not the only route to reducing dissonance proposed in the original theory by Festinger (1957), this is the route that is generally equated with dissonance reduction in the literature on cognitive dissonance. So, according to these studies, participants in the Adriaanse et al. (2014) experiment who indulged in chocolates as a result of hedonic priming would restore consistency not by confabulating a reason that justifies the behavior, but rather by changing their attitudes or standards about dieting (e.g., by decreasing the importance of dissonant cognitions; 'Dieting is not the most important thing there is in life'). Another difference is that in dissonance studies people may experience insufficient justification for their behavior (e.g., in forced compliance paradigms such as Festinger & Carlsmith, 1959; Beauvois, Bungert, & Mariette, 1995) but in the present studies, people experience a complete lack of justification for their behavior as they are unaware of the cause of their behavior. It makes sense to assume that in the latter case misattribution processes, like attitude change or confabulation, are even more likely to occur. Future research is necessary to integrate these lines of research and investigate in which situations people are more likely to change their standards (i.e., use classical dissonance reduction) or keep their standards intact and resort to confabulation when attempting to restore consistency.

Long-term effects of confabulation

Although our tendency to construct post-hoc reasons for unexpected behavior when the actual reasons are inaccessible is probably adaptive in terms of allowing people to make sense of their world, it is not difficult to imagine that this can – depending on the content of the explanations – also have negative consequences for future self-control attempts. Indeed, there have been scholars arguing that for long-term self-control success, regulating responses to failure might be crucial, and arguably even more important than preventing single instances of failure. For example, Baumeister and Heatherton (1996) already argued that "although considerable research has focused on what causes people to violate their standards or other self-regulatory patterns, it is important to realize that the majority of such violations are inherently trivial" (p. 11). Surely, skipping one night of your exercise regime to watch your favorite TV show hardly affects your long-term health goals or threatens your fitness level. Rather than this one violation in itself, it is the subsequent chain of behavior which may be set into motion by, amongst other things, people's persistent need to explain their behavior to themselves which is detrimental to long-term self-control success (Baumeister & Heatherton, 1996). Coming back to our example of Anna, if

she attributes her indulgent behavior to feelings of stress, this may result in perceiving herself as an emotional eater, which, in turn, may become a self-fulfilling prophecy the next time Anna feels stressed and is presented with tempting foods.

Although evidence for the suggestion that spontaneous confabulations can have long-term effects and turn into self-fulfilling prophecies is still missing, research on provoked confabulation has provided support for the notion that the way we deal with norm-violating behavior may impact future behavior beyond this one instance of failure: Bar-Anan et al. (2010) provided initial evidence that in the case of provoked confabulation, confabulated reasons can indeed become integrated into self-knowledge and affect subsequent behavior. In addition, there are several studies that have provided support for the broader notion that misattributed internal states get incorporated into people's self-concepts which in turn affect future behavior (e.g., Fazio, Effrein, & Falender, 1981), making the self-fulfilling prophecy effect of spontaneous confabulation a plausible chain of effects which warrants further scrutiny in future research.

Summary

There are many situations where we do not have access to the causes of our behavior (Nisbett & Wilson, 1977; Sheeran et al., 2013). In those cases where we are confronted with behavior that signals self-control failure (i.e., behavior that does not align with our long-term goals or standards) and for which we do not have an explanation, we experience elevated levels of negative affect. To reduce this negative affect, people may spontaneously confabulate plausible reasons for acting. Future research is needed to test the degree to which this process occurs reflexively and spontaneously and to examine the implications for the formation of self-knowledge as well as for future self-control attempts.

References

Aarts, H., Custers, R., & Wegner, D. M. (2005). On the inference of personal authorship: Enhancing experienced agency by priming information. *Consciousness and Cognition: An International Journal*, *14*, 439–445.

Adriaanse, M. A., Kroese, F. M., Weijers, J., Gollwitzer, P. M., & Oettingen, G. (in press). Explaining unexplainable food choices. *European Journal of Experimental Psychology*.

Adriaanse, M. A., Prinsen, S., De Witt Huberts, J. C., De Ridder, D. T. D., & Evers, C. (2016). 'I ate too much so I must have been sad': Emotions as a confabulated reason for overeating. *Appetite*, *103*, 318–323.

Adriaanse, M. A., Weijers, J., De Ridder, D. T. D., De Witt Huberts, J. C., & 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 behavior. *Journal of Experimental Social Psychology*, *46*, 884–894.

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.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, *74*, 1252–1265.

Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological Inquiry*, *7*, 1–15.

Beauvois, J. L., Bungert, M., & Mariette, P. (1995). Forced compliance: Commitment to compliance and commitment to activity. *European Journal of Social Psychology*, *25*, 17–26.

Blanken, I., Van de Ven, N., & Zeelenberg, M. (2015). A meta-analytic review of moral licensing. *Personality and Social Psychology Bulletin*, *41*, 540–558.

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

Cooper, J., & Fazio, R. H. (1984). A new look at dissonance theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 229–262). Hillsdale, NJ: Erlbaum.
De Witt Huberts, J. C., Evers, C., & De Ridder, D. T. D. (2012). License to sin: Self-licensing as a mechanism underlying hedonic consumption. *European Journal of Social Psychology*, *42*, 490–496.
De Witt Huberts, J. C., Evers, C., & De Ridder, D. T. D. (2014). 'Because I am worth it': A theoretical framework and empirical review of a justification-based account of self-regulation failure. *Personality and Social Psychology Review*, *18*, 119–138.
Effron, D. A., Monin, B., & Miller, D. T. (2013). The unhealthy road not taken: Licensing indulgence by exaggerating counterfactual sins. *Journal of Experimental Social Psychology*, *49*, 573–578.
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.
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*. Stanford, CA: Stanford University Press.
Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, *58*, 203–211.
Fishbach, A., & Dhar, R. (2005). Goals as excuses or guides: The liberating effect of perceived goal progress on choice. *Journal of Consumer Research*, *32*, 370–377.
Fujita, K. (2011). On conceptualizing self-control as more than the effortful inhibition of impulses. *Personality and Social Psychology Review*, *15*, 352–365.
Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication. *Brain*, *123*, 1293–1326.
Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, *136*, 495–525.
Harmon-Jones, E., Harmon-Jones, C., & Levy, N. (2015). An action-based model of cognitive dissonance processes. *Current Directions in Psychological Science*, *24*, 184–189.
Harris, J. L., Bargh, J. A., & Brownell, K. D. (2009). Priming effects of television food advertising on eating behavior. *Health Psychology*, *28*, 404–413.
Hirstein, W. (2009). *Confabulation: Views from neuroscience, psychiatry, psychology and philosophy*. Oxford: Oxford University Press.
Hofmann, W., Friese, M., & Wiers, R. W. (2008). Impulsive versus reflective influences on health behavior: A theoretical framework and empirical review. *Health Psychology Review*, *2*, 111–137.
Jordan, J., Mullen, E., & Murnighan, J. K. (2011). Striving for the moral self: The effects of recalling past moral actions on future moral behavior. *Personality and Social Psychology Bulletin*, *37*, 701–713.
Khan, U., & Dhar, R. (2006). Licensing effect in consumer choice. *Journal of Marketing Research*, *153*, 259–266.
Kivetz, R., & Zheng, Y. (2006). Determinants of justification and self-control. *Journal of Experimental Psychology: General*, *135*, 572–587.
Kopelman, M. D. (1987). Two types of confabulation. *Journal of Neurology, Neurosurgery & Psychiatry*, *50*, 1482–1487.
Mazar, N., & Zhong, C. B. (2010). Do green products make us better people? *Psychological Science*, *21*, 494–498.
Molden, D. C. (2014). Understanding priming effects: What is 'social priming' and how does it occur? In D. C. Molden (Ed.), *Understanding priming effects in social psychology* (pp. 3–14). New York: The Guilford Press.
Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, *81*, 33–43.
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). Nonconscious goal pursuit: Acting in an explanatory vacuum. *Journal of Experimental Social Psychology*, *42*, 668–675.
Parks-Stamm, E. J., Oettingen, G., & Gollwitzer, P. M. (2010). Making sense of one's actions in an explanatory vacuum: The interpretation of nonconscious goal striving. *Journal of Experimental Social Psychology*, *46*, 531–542.
Prinsen, S., Evers, C., & De Ridder, D. T. D. (2016). Oops I did it again: Examining self-licensing effects in a subsequent self-regulation dilemma. *Applied Psychology: Health and Well-Being*, *8*, 104–126.
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.

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.

Taylor, C., Webb, T. L., & Sheeran, P. (2013). 'I deserve a treat!': Justifications for indulgence undermine the translation of intentions into action. *British Journal of Social Psychology*, *53*, 501–520.

Tversky, A., & Kahneman, D. (1974). Judgements under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

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

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

Weibel, C., Messner, C., & Brügger, A. (2014). Completed egoism and intended altruism boost healthy food choices. *Appetite*, 77, 38–45.

Wheatley, T. (2009). Everyday confabulation. In B. Hirstein (Ed.), *Confabulation: Views from neuroscience, psychiatry, psychology, and philosophy* (pp. 205–225). Oxford: Oxford University Press.

Wilcox, K., Kramer, T., & Sen, S. (2011). Indulgence or self-control: A dual process model of the effect of incidental pride on indulgent choice. *Journal of Consumer Research*, *38*, 151–163.

Zemack-Rugar, Y., Corus, C., & Brinberg, D. (2012). The 'Response-to-failure' scale: Predicting behavior following initial self-control failure. *Journal of Marketing Research*, *49*, 996–1014.