



Universiteit
Leiden
The Netherlands

Understanding the setback effect in everyday self-regulation

Broeke, P. ten; Adriaanse, M.A.

Citation

Broeke, P. ten, & Adriaanse, M. A. (2023). Understanding the setback effect in everyday self-regulation. *European Journal Of Social Psychology*. doi:10.1002/ejsp.2931

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3627586>

Note: To cite this publication please use the final published version (if applicable).

RESEARCH ARTICLE

Understanding the setback effect in everyday self-regulation

Pam ten Broeke¹ | Marieke A. Adriaanse²¹Behavioural Science Institute, Radboud University, Nijmegen, Netherlands²Department of Social, Health and Organizational Psychology, Utrecht University, Utrecht, Netherlands

Correspondence

Pam ten Broeke, Behavioural Science Institute, Radboud University, Nijmegen, Netherlands.
Email: pam.tenbroeke@ru.nl

Pam ten Broeke and Marieke A. Adriaanse shared first authorship.

Marieke A. Adriaanse is now at the Department of Health, Medical and Neuropsychology & Leiden University Medical Center, Leiden University, Leiden, Netherlands.

Abstract

People experience minor instances of self-regulation failure, i.e., setbacks. While in itself trivial, a single setback may trigger a maladaptive chain of cognitions and emotions leading to more self-regulation failure. In the present article, we investigated (a) whether the experience of a setback indeed increases the chance of subsequent failure, and (b) what psychological mechanisms underly this effect. We conducted three studies in the context of eating behaviour, exercising, and saving. At T1 we assessed participants intentions, self-efficacy, and perceived usefulness to self-regulate in response to a recent setback (versus an instance of self-regulation success). At T2 we assessed subsequent self-regulation failure and success using retrospective recall. Experiencing a setback increased the likelihood of subsequent failure and decreased the chances of subsequent success. Our results most consistently highlight a decrease in self-efficacy as an underlying mechanism, and therefore as a promising target for interventions to prevent the setback effect from occurring.

KEYWORDS

disinhibition, intentions, self-efficacy, self-regulation failure, what the hell

1 | INTRODUCTION

Self-regulation—the regulation of one's own behaviour to achieve a certain desired end state (Carver & Scheier, 1998; Inzlicht et al., 2021)—is a crucial skill in life: self-regulation skills, such as the ability to set goals, deal with temptations and monitor progress, are needed to eat healthily, exercise, pay bills, say no to drugs, and study for exams despite being tempted to go out with friends. Failure to self-regulate, on the contrary, is assumed to be at the core of many societal problems, including obesity, substance abuse, dropping out of school, financial problems, and many more (Ridder et al., 2017). To prevent such self-regulation failure, psychologists have dedicated decades of research to understand the precursors of self-regulation failure and to develop helpful strategies and interventions to prevent self-regulation failure from occurring (Michie et al., 2014). Yet, even while using helpful strategies, people still experience many barriers to self-regulation on a

daily basis, such as conflicting goals, and inherent limits in resources such as attention, memory, and motivation (e.g., Milyavskaya & Inzlicht, 2017). As a consequence, minor instances of failure—that is, setbacks (Wenzel et al., 2020)—such as yelling at someone despite meaning to be calm, or overspending despite trying to save money, remain common experiences in our everyday lives (Adriaanse & ten Broeke, 2022).

Crucially, while in themselves most setbacks are inherently trivial (Baumeister & Heatherton, 1996), they may lead to a chain of more and more self-regulation failures. For example, eating one piece of cake on your grandmother's birthday on its own hardly threatens your long-term goal to lose 5 pounds. Yet, rather than this setback itself, the subsequent chain of thoughts, emotions, and behaviours which may be set in motion upon experiencing a setback can be detrimental to long-term self-regulation success (Baumeister & Heatherton, 1996; Wagner & Heatherton, 2015). Indeed, restrictive dieters who were asked to violate their diets by consuming a high-caloric milkshake, voluntarily ate

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *European Journal of Social Psychology* published by John Wiley & Sons Ltd.

more calories in a subsequent alleged 'taste test' compared to restrictive dieters who had not violated their diets before, and compared to non-dieters (Herman & Mack, 1975). Also, when a person suffering from addiction disorder experiences a single relapse in abstinence behaviour, they are more likely to fully relapse into their addiction behaviour, especially when the person believes the reasons for the initial lapse are within themselves and uncontrollable (e.g., Curry et al., 1987). In sum, research within the context of restrictive dieting and addiction suggests that a single instance of failure may set into motion subsequent failures and will hamper future self-regulation success.

A recent study investigated the setback effect in the context of everyday, non-pathological, self-regulation behaviour. Participants in this study reported that, opposed to the common belief that self-control success will be lowest after exerting self-control (i.e., the classic ego depletion effect; Baumeister et al., 1998), they experienced the lowest self-control success rates after experiencing a setback, that is, when prior self-control efforts failed (Wenzel et al., 2020). To our knowledge, this is one of the few studies that have focused on the setback effect in the context of what we call 'everyday self-regulation' (as opposed to self-regulation in the context of pathological behaviour such as addiction or restrictive dieting). Moreover, the study by Wenzel et al. (2020) is within the field of self-regulation one of the few studies to investigate what happens *after* people have experienced a setback in everyday self-regulation. Overall, the majority of research endeavours in the field of self-regulation has focused on understanding predictors of setbacks, and developing strategies to prevent setbacks from occurring. As a consequence, even though it seems inevitable for people to experience setbacks in healthy eating, exercising, going to bed on time, and many more everyday self-regulation behaviours (Adriaanse & ten Broeke, 2022), psychologists have few insights into what happens in the aftermath of such setbacks.

Considering the potential negative downstream effects of initially relatively harmless setbacks, it is important also to develop strategies to better prepare people for setbacks in order to prevent them from snowballing into a downward spiral of failure. However, in order to achieve this, we need more insights into whether and how setback effects occur in everyday self-regulation. The aim of this study is therefore to investigate (a) whether we can replicate the finding of Wenzel et al. (2020) that the experience of a setback increases the chance of subsequent failure in the context of everyday self-regulation, and (b) what psychological mechanisms underly this detrimental effect. Insights into the existence of a setback effect in everyday self-regulation will help determine the need for strategies to help people cope with setbacks. Insights into the underlying mechanisms of a setback effect will provide concrete starting points for the development of such strategies. We explore three potential routes through which a single setback could result in subsequent failure: Do people no longer *want* to self-regulate? Do people no longer feel *capable* to self-regulate? Or do people perceive future self-regulation to be *useless*?

First, after a setback, people may experience a decrease in their *intention* to self-regulate. Research on Cognitive Dissonance Theory (Stone & Cooper, 2001) shows that motivational constructs are a key source of misattribution when people are confronted with behaviours

that are inconsistent with their long-term goals (e.g., 'Apparently, I am not that motivated to adhere to my diet'). Moreover, intentions are an essential predictor of behaviour according to many social-cognitive theories (e.g., Ajzen, 2002). So, when experiencing a setback in self-regulation, we expect people to experience a decrease in intentions to self-regulate, which in turn increases the chance of subsequent self-regulation failure.

Second, after a setback, people may experience a decreased confidence in their ability to self-regulate, that is, a decrease in *self-efficacy*. People essentially derive their self-efficacy from their performance in previous situations (Bandura, 1978). Indeed, decreased feelings of personal control play a key role in the abstinence violation effect (Marlatt & Witkiewitz, 2005): after experiencing a lapse, addicts typically feel less personal control, which increases the chance of a full-blown relapse into the addiction behaviour. Moreover, previous work has confirmed that low self-efficacy consistently predicts relapse into physical inactivity and unhealthy eating (Roordink et al., 2021). So, when experiencing a setback in self-regulation, we expect people to experience a decrease in self-efficacy, which in turn increases the chance of subsequent self-regulation failure.

Third, after a setback, people may experience a decreased *perceived usefulness* to self-regulate. This idea is in line with the so-called 'what the hell effect' that is suggested to underly the classic finding that restrictive dieters overeat after an initial diet violation (Herman & Mack, 1975). That is, researchers assume that restrictive dieters have an 'all-or-nothing' mentality: they perceive a single diet violation to ruin the diet completely, and therefore any future attempts to adhere to the diet have become useless (Herman & Mack, 1975; Wagner & Heather-ton, 2015). However, researchers have experienced difficulties with pinpointing the proposed cognitive underlying mechanism (Jansen et al., 1988). Moreover, more recent studies did not consistently show evidence for the disinhibition effect (e.g., Jansen et al., 2009; Tomiyama et al., 2009). It thus remains to be determined to which extent a decreased perceived usefulness to self-regulate may be responsible for fuelling the setback-effect of self-regulation failure, in particular in the context of everyday self-regulation. Therefore, we test the hypothesis that when experiencing a setback in self-regulation, people experience a decrease in the perceived usefulness to self-regulate, which in turn increases the chance of subsequent self-regulation failure.

2 | THE PRESENT STUDIES

We conducted three studies all of which involved the context of eating behaviour, as reducing unhealthy food consumption represents a prototypical, frequently studied 'everyday' self-regulation dilemma which is relevant to many people. In addition, considering the high number of food decisions people encounter every day, it is likely that most people trying to restrict their unhealthy food intake will experience minor diet violations on a relatively frequent basis, which creates a relevant context for testing our hypotheses. In order to demonstrate the generalizability of our findings across behavioural domains, Study 3 included three additional self-regulation

domains that people frequently pursue, and in which many people regularly experience instances of failure, namely exercising and saving.

Across the three studies, setbacks were measured in real-life contexts. We specifically chose not to manipulate actual failure versus success (as is for example done in preload paradigms classically used in the literature on the disinhibition effect; Herman et al., 1987) as this could induce unwanted side effects, such as providing participants with a clear situational explanation for the failure (e.g., 'because I had to for the experiment') and it would limit our ability to draw meaningful conclusions regarding the relevance and working mechanisms of the setback effect in everyday self-regulation. Moreover, simulating a sequence of self-regulation dilemmas in a controlled laboratory setting would hardly reflect the constant flow of self-regulation behaviour and accompanied responses in real-life (cf. Tomiyama et al., 2009).

Statistical analyses were performed in R (Version 4.1.2). For information about handling of assumptions and outliers, please refer to the Appendix. For materials of all studies, please refer to the supplementary files.¹ Unless otherwise indicated, statistical assumptions were met, or the statistical test was expected to be robust against violations.

3 | STUDY 1

In Study 1 we aimed to get a first glimpse of potential changes in intention, self-efficacy, and perceived usefulness after a setback in self-regulation, and subsequent self-regulation failure in the context of dieting. As such we looked at self-regulation failure (i.e., setbacks) only. Intention, self-efficacy, and perceived usefulness were assessed retrospectively, and participants were asked to think back to their most recent setback. Subsequent failure was assessed as the frequency of diet violations and average dieting success in the four days after our assessment of intention, self-efficacy, and perceived usefulness, through retrospective recall.

3.1 | Method

3.1.1 | Participants

We recruited through Prolific a total of 225 female, native English-speaking participants, aged 18–40, currently residing in the United Kingdom. Based on post-hoc sensitivity calculations in G*Power (Faul et al., 2007), this sample size ensured a power of 0.90 ($\alpha = 0.05$) for detecting an estimated partial $R^2 = .05$ (small to medium effect), when regressing the outcome measures on self-efficacy or perceived usefulness, including three covariates, and accounting for a drop-out of 10%. We included only participants who were currently restricting their

food intake to manage their weight and who recently violated their diets (see Materials). Seven participants dropped out. The final sample included 218 participants with an average age of 30.87 ($SD = 5.45$) and an average BMI of 28.01 ($SD = 7.36$).

3.1.2 | Design and procedure

We used a longitudinal design with changes in intention, self-efficacy, and perceived usefulness after failure assessed at T1. At T2, we assessed the dependent variables failure T2 and success T2, aggregated over four subsequent days.² As age, BMI, and dietary restraint (Herman & Mack, 1975) have been often related to dieting behaviour in previous research, we assessed these as potential confounders.

After informed consent, demographics and dietary restraint were assessed. Next, we assessed whether participants violated their diet in the past 4 days. Only participants answering 'yes' completed the rest of the study. We assessed the time since the setback and the degree of the setback as potential confounders.

Participants recalled the most recent situation in which they violated their diet. Next, participants answered questions about their relative ('as compared to before the setback') intention, self-efficacy, and perceived usefulness after experiencing the most recent setback. At 7 a.m. (local time) 5 days after part 1, participants were invited to complete part 2 within 39 h (until 10 p.m. the next day). Participants indicated their instances of failure (failure T2) and perception of success (success T2).³ The study ended with a debriefing and reimbursement of £4.50.

3.1.3 | Materials

Dietary restraint

Dietary restraint was assessed using the Restraint Scale (RS; Herman & Polivy et al., 1980), consisting of 10 items, which were summed and answered on a 4- or 5-point scale (e.g., 'How often are you dieting?' with response options: 'never', 'rarely', 'sometimes', 'usually', 'always'; Cronbach's $\alpha = .73$). Higher scores indicate more dietary restraint.

² Two additional dependent variables were assessed (and preregistered) in Study 1 and Study 2, namely the total number of days in which participants failed over the chosen time period (*failure days*; Study 1 and 2), and participants' perceptions of *overall success* over the total time period (Study 1). Failure days were highly correlated with failure T2, and all results were the same as for failure T2 except for the mediation effect through self-efficacy and perceived usefulness in Study 2 and the mediation effect through perceived usefulness in the eating domain in Study 3 (both were significant for failure T2 but not for failure days T2). Moreover, the MANOVAs that would originally include failure days T2 according to the preregistration of Study 2 and Study 3 showed similar results when failure days T2 was included versus excluded. The variable *overall success* correlated highly with success T2 and showed similar results. Therefore, for the sake of brevity, in this article we only reported results on failure T2 and success T2.

³ In all studies participants additionally responded to some questions on the extent to which they were able to recall their dieting, exercising, saving and/or procrastination behaviour of the respective day. Average recall was high in all studies ($M's \geq 5.38$ on a 7-point scale). For all studies, additional analyses revealed that excluding participants with low average recall (<4) did not change the pattern of results, except that the indirect effect of regulation on failure T2 and success T2 through perceived usefulness was no longer significant in the eating domain in Study 3. Exact questions can be found in the Supplementary Materials.

¹ In all studies, participants responded to several additional questionnaires which are outside the scope of the present article. A complete overview of all used materials for all studies can be found in the Supplementary Materials.

Recent diet violation

Participants were asked: 'In the past 4 days, did you eat something—a meal or a snack—that was not in line with your dieting goal?' Participants responding 'No' (6%) were excluded. Participants responding 'Yes' received open-ended questions about the social and emotional context of the situation to stimulate vivid recollection (Eisenhower et al., 1991). As potential covariates, participants indicated the degree to which they violated their diet (failure degree) on a scale ranging from 1 (*very small violation*) to 7 (*very large violation*), and how many days ago the situation took place (time since failure). To stimulate accurate recollection, participants were provided with the image of a calendar (Eisenhower et al., 1991).

Intention, self-efficacy, perceived usefulness

Intention, self-efficacy, and perceived usefulness after the diet violation were assessed using a 'compared to before' format (cf. Adriaanse et al., 2010). Intention was measured by one item:⁴ 'Compared to before you violated your dieting goal, to what extent did you intend to restrict your food intake after you violated your dieting goal?' (Adriaanse et al., 2010). Self-efficacy was measured by two items ($r = .74$; Ajzen, 2002), e.g., 'Compared to before you violated your dieting goal, how confident did you feel about your abilities to restrict your food intake after you violated your dieting goal?' Perceived usefulness was assessed by two items ($r = 0.77$), e.g., 'Compared to before you violated your dieting goal, to what extent did you feel it was worthwhile to restrict your food intake from that moment on after you violated your dieting goal?' Variables were answered on Likert scales ranging from 1 (*much less*) to 7 (*much more*). As such, response options above the midpoint (4) of the scale indicated an *increase* and response options below the midpoint indicated a *decrease* in intentions, self-efficacy, or perceived usefulness compared to before the diet violation.

Failure and success at T2

Failure T2 was the sum of four retrospective daily ratings of failure frequency ('How many times during this day did you eat something that was not in line with your dieting goal?'). *Success T2* was the mean of four retrospective daily ratings of dieting success ('To what extent do you feel like you successfully followed your diet during this day?') answered on a Likert scale ranging from 1 (*not at all*) to 7 (*very much*).

3.1.4 | Data-analysis

To assess whether participants reported a significant *increase* (>4) or *decrease* (<4) in intention, self-efficacy and perceived-usefulness, we conducted one-sample t-tests compared the mean to $\mu = 4$. Next, we examined intercorrelations to assess whether the potential covariates (age, BMI, dietary restraint, time since failure, and degree of failure)

were significantly related to the outcome measures, and therefore needed to be included as covariates.

To investigate whether relative intention, self-efficacy, and/or perceived usefulness predicted failure T2 and success T2, we conducted hierarchical regression analyses. In step 1, we included relevant covariates. In step 2, we added intention, self-efficacy, or perceived usefulness.

3.2 | Results

3.2.1 | Descriptives and intercorrelations

See Table 1 for descriptives and intercorrelations. Surprisingly, after violating their diets, participants reported on average an *increase* (>4) in intention ($M = 5.33$, $SD = 1.41$, $p < .001$) and perceived usefulness ($M = 4.53$, $SD = 1.69$; $p < .001$). As expected, participants on average reported a *decrease* (<4) in self-efficacy ($M = 3.70$, $SD = 1.49$; $p = .004$). See also the distributions of intention, self-efficacy, and perceived usefulness in Figure 1.

Intention was not related to the outcome measures, $ps \geq .112$. Self-efficacy and perceived usefulness were both significantly correlated with success T2, $ps \leq .008$, but not associated with failure T2, $ps \geq .144$.

3.2.2 | Main analyses

Because dietary restraint and time since failure were significantly related to success T2, these were included as covariates. Due to technical issues, one participant had a missing response on perceived usefulness, leaving $N = 217$ for the analyses including perceived usefulness. Because assumptions of normality were violated for time since failure, we used bias-corrected bootstrapped 95% confidence intervals. We rejected the null hypothesis when the confidence interval excluded zero (See Table 2 for statistics). Step 1 was significant for both outcome measures ($ps < .002$), with dietary restraint, but not time since failure, generally predicting more failure and less success at T2.

Adding intention did not significantly improve the model for failure T2 ($p = .507$) and success T2 ($p = .799$). Adding self-efficacy did not significantly improve the model for failure T2 ($p = .231$), but significantly improved the prediction of success T2 ($p = .009$), with higher self-efficacy predicting more success at T2 ($B = 0.14$, 95% CI [0.03, 0.26]; see Figure 1). Adding perceived usefulness did not significantly improve the model for failure T2 ($p = .281$), but significantly improved the prediction of success T2 ($p = .003$), with higher perceived usefulness predicting more success at T2 ($B = 0.14$, 95% CI [0.04, 0.24]; see Figure 1).

3.3 | Discussion

Results from Study 1 indicated that when comparing how people felt after failure to before failure, on average, people reported that their

⁴ In Studies 1 and 2, intention was originally measured with two items. The additional item was 'Compared to before you violated your dieting goal, how motivated did you feel to restrict your food intake after you violated your dieting goal?'. However, responses on this item correlated poorly with the 'intention' item, and showed an unexpected correlational structure with other items. Therefore, we decided to only include the *intention* item for this construct.

TABLE 1 Study 1: means (M), standard deviations (SD), and intercorrelations of key variables

	1	2	3	4	5	6	7	8	9	10
Age (1)										
BMI (2)	0.20									
Dietary restraint (3)	−.04	.38								
Time since failure (4)	.03	0.09	−0.08							
Degree of failure (5)	.01	.21	.32	.04						
Intention (6)	.10	.06	.30	.00	.19					
Self-efficacy (7)	0.09	.02	−.06	.10	−.09	.24				
Perceived usefulness (8)	.05	−.09	−.03	−.14	−.05	.28	.41			
Failure T2 (9)	−.11	.10	.22	−.09	.04	.11	−0.10	−0.07		
Success T2 (10)	.08	−0.03	−.21	0.16	−0.07	−.04	0.20	0.18	−0.73	
M	30.87	28.01	20.00	1.85	5.32	5.33	3.70	4.53	4.18	41.64
SD	5.45	7.36	4.81	1.09	1.52	1.41	1.49	1.69	2.98	1.21

Note: Statistically significant correlation coefficients ($p < .05$) are shown in bold text.

TABLE 2 Study 1: model statistics, regression coefficients and bias-corrected bootstrapped 95% confidence intervals (CI) of the regressions including intention, self-efficacy, or perceived usefulness on failure T2 and success T2

	Failure T2			Success T2		
	B	95% CI	ΔR^2	B	95% CI	ΔR^2
Intention						
Step 1			.06**			.06***
Time since failure	−0.19	[−0.51, 0.20]		0.16	[0.00, 0.30]	
Dietary restraint	0.14	[0.05, 0.23]		−0.05	[−0.08, −0.02]	
Step 2						
Time since failure	−0.20	[−0.52, 0.19]	<.001	0.16	[−0.00, 0.30]	<.001
Dietary restraint	0.13	[0.04, 0.23]		−0.05	[−0.08, −0.02]	
Intention	0.10	[−0.19, 0.36]		0.02	[−0.10, 0.13]	
Self-efficacy						
Step 1			.06**			.06***
Time since failure	−0.19	[−0.51, 0.20]		0.16	[0.00, 0.30]	
Dietary restraint	0.14	[0.05, 0.23]		−0.05	[−0.08, −0.02]	
Step 2			.01			.03**
Time since failure	−0.17	[−0.50, 0.22]		0.14	[−0.02, 0.28]	
Dietary restraint	0.13	[0.05, 0.23]		−0.05	[−0.07, −0.02]	
Self-efficacy	−0.16	[−0.43, 0.07]		0.14	[0.03, 0.26]	
Perceived usefulness						
Step 1			.05**			.06***
Time since failure	−0.19	[−0.54, 0.22]		0.16	[−0.01, 0.31]	
Dietary restraint	0.14	[0.05, 0.23]		−0.05	[−0.08, −0.02]	
Step 2			0.01			.04**
Time since failure	−0.21	[−0.54, 0.14]		0.19	[0.04, 0.33]	
Dietary restraint	0.13	[0.05, 0.24]		−0.05	[−0.08, −0.02]	
Perceived usefulness	−0.13	[−0.36, 0.18]		0.14	[0.04, 0.24]	

Note: Statistically significant confidence intervals (excluding zero) are shown in bold text.

** $p < .01$, *** $p < .001$.

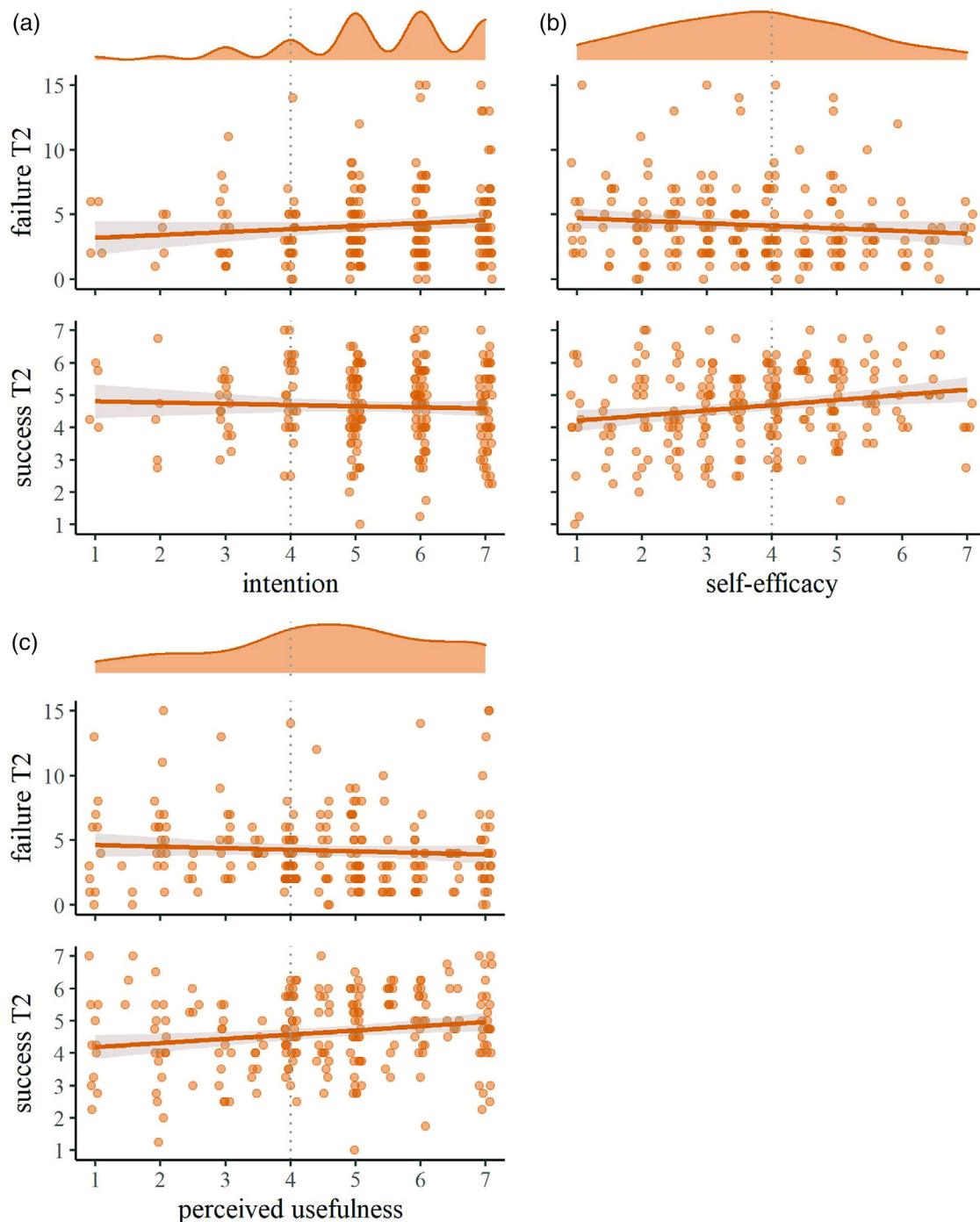


FIGURE 1 Study 1: density distributions of intention (a), self-efficacy (b), and perceived usefulness (c), and their relationship to failure T2 and success T2. Grey areas reflect 95% confidence intervals. The grey dotted line shows the midpoint (4) of the scale.

intentions and the perceived usefulness to pursue their dieting goal increased, while their confidence in their abilities to do so dropped. The change in self-efficacy as well as the change in perceived usefulness after experiencing failure predicted subsequent success, but not failure. Changes in intention did not predict subsequent dieting failure or success. Together these results tentatively point to the relevance of self-efficacy and, to a lesser extent, perceived usefulness in fuelling a setback effect.

4 | STUDY 2

Study 2 was conducted as a replication and extension of Study 1. That is, in Study 2 we again tested the role of intention, self-efficacy, and perceived usefulness, but this time we compared participants who recently violated their diet to participants who were successful in adhering to their diet to allow for a full test of the mediation model. The

study was preregistered at the open science framework: https://osf.io/2p85e/?view_only=0590af0b7f044e26ad16e3d9aa3ce8db. The main effect of previous regulation (success or failure) on subsequent failure and success was registered as confirmatory as it has been previously reported that experiencing failure (a setback) has the potential of snowballing into subsequent failure (Wenzel et al., 2020). The mediating role of self-efficacy in fuelling this snowball effect was also registered as confirmatory based on our findings in Study 1 as well as recent work (Adriaanse & ten Broeke, 2022; Roordink et al., 2021). All other effects were labelled exploratory, including the roles of intention and perceived usefulness, as these unexpectedly increased instead of decreased upon experiencing failure in Study 1.

4.1 | Method

4.1.1 | Participants

We recruited through Prolific a total of 208 female, native English-speaking participants, aged 18–40, currently residing in the United Kingdom, and who had not participated in Study 1. This sample size was based on a-priori power analyses for all confirmatory hypotheses in G*Power and Glimpse, the largest required sample size being $n = 238$ for regressing total failure or success T2 on self-efficacy, including three potential covariates, aiming for a power of 0.90 ($\alpha = .05$) and a small effect ($R^2 = .05$, estimated based on Study 1), accounting for a drop-out of 15%. Unexpectedly, a higher number of participants (23%) dropped out between part 1 and part 2, leaving a final sample of 158 participants with an average age of 30.22 ($SD = 6.12$) and an average BMI of 29.27 ($SD = 7.28$). Participants who dropped out were equally divided over the two regulation conditions, $p = .880$, and did not significantly differ from participants in the final sample in age, BMI, dietary restraint, self-efficacy, intention, or perceived usefulness, $ps \geq 0.205$.

4.1.2 | Design, procedure and materials

Study 2 followed a similar procedure as Study 1 with a few exceptions. We assessed regulation by asking participants whether they, today or yesterday, ate something that was not in line with their dieting goal. We changed the time window from 4 days to 2 days, as based on data on the 'time since failure' question from Study 1 this was expected to yield a group responding 'yes' (failure condition) that would be approximately equal to a group responding 'no' (success condition). As we reduced the time-window, we also no longer assessed 'time since failure' as potential covariate. After answering several questions about their diet, participants were asked to recall the most recent situation in which they had violated their diet (failure condition) or had resisted violating their diet (success condition). Next, they answered the questions on self-efficacy ($r = .80$), intention, and perceived usefulness ($r = .77$) relative to before their most recent failure or success (framed as 'after you violated your diet' for the failure condition, and as 'after you successfully adhered to your diet' in the success condition). At T2, we assessed

the same variables as in Study 1. Participants had 15 h (until 10 p.m. the same day) to complete part 2, and they received £3.70 for participating. Reliability of the RS was adequate (Cronbach's $\alpha = .77$).

4.1.3 | Data-analysis

We examined intercorrelations to assess whether age, BMI and, dietary restraint were significantly related to the outcome measures, and therefore needed to be included as covariates. To test the effect of regulation (failure vs. success) on failure T2 and success T2, and on intention, self-efficacy, and perceived usefulness, two Multivariate ANOVAs (MANOVAs) were performed, followed by one-way ANOVAs. To assess whether participants reported a significant *increase* (>4) or *decrease* (<4) in intention, self-efficacy, and perceived-usefulness, we conducted one-sample t-tests compared the mean to $\mu = 4$ within the failure condition and the success condition separately. In case of a significant association between regulation and the mediators (direct effect), and between the mediators and the outcome measures, we performed bias-corrected bootstrapped tests of the indirect effect using the 'mediate' package in R with 1000 bootstrapped samples.

4.2 | Results

4.2.1 | Descriptives and intercorrelations

Descriptives and intercorrelations of key variables are depicted in Table 3. An unexpectedly large proportion of the participants ($n = 128$) were in the failure condition and only 30 participants were in the success condition. Intention, self-efficacy, and perceived usefulness were significantly and negatively related to failure T2 ($ps \leq .006$) and positively related to success T2 ($ps \leq .012$). The results of independent t-tests indicated that age, BMI, and dietary restraint were not different between conditions, $ps \geq .243$. As dietary restraint was (unlike Study 1) not related to the outcome measures, this variable was, in contrast to what was preregistered, not included as covariate in the analyses.⁵

4.2.2 | Main analyses

We performed a square root transformation on failure T2 to normalize the distribution. To facilitate interpretation, means and standard deviations are presented for the untransformed variable. The results are presented in Table 4 and Figure 2.

Participants reported significantly more failure and lower success at T2 after initial failure compared to after initial success. Participants reported an increase (>4) in intention both after initial failure ($p < .001$) and after initial success ($p < .001$), but relative intention was lower

⁵ Sensitivity analyses indicated that including dietary restraint as a covariate did not change the pattern of results.

TABLE 3 Study 2: means (M), standard deviations (SD), and intercorrelations

	1	2	3	4	5	6	7	8
Age (1)								
BMI (2)	.25							
Dietary restraint (3)	−.01	0.39						
Intention (4)	.00	−.05	.18					
Self-efficacy (5)	−.06	0.13	−.09	0.32				
Perceived usefulness (6)	.04	−0.14	−.18	.31	0.56			
Failure T2 (7)	0.04	0.05	0.06	−0.22	−0.23	−0.22		
Success T2 (8)	0.02	−.15	−.10	.20	0.30	0.29	−0.70	
M	30.22	29.27	20.82	5.05	4.10	4.31	4.85	4.94
SD	6.12	7.28	5.09	1.47	1.71	1.81	4.59	1.33

Note: Statistically significant correlation coefficients ($p < .05$) are shown in bold text.

TABLE 4 Study 2: means (M), standard deviations (SD), and statistics of the MANOVAs and follow-up ANOVAs on the effect of regulation on failure T2 and success T2, and on intention, self-efficacy and perceived usefulness

	Regulation T1		(M)ANOVA results			
	Failure group (N = 128)	Success group (N = 30)				
	M (SD)	M (SD)	F	df1, df2	p	partial η^2
Multivariate			5.00	2, 155	.008	
Failure T2	5.27 (4.78)	3.10 (3.14)	9.95	1, 156	.002	0.06
Success T2	4.82 (1.34)	5.41 (1.19)	4.84	1, 156	.029	0.03
Multivariate			11.53	3, 154	<.001	
Intention	4.91 (1.50)	5.63 (1.16)	6.03	1, 156	.015	0.04
Self-efficacy	3.79 (1.64)	5.43 (1.33)	26.04	1, 156	< .001	0.14
Perceived usefulness	3.99 (1.80)	5.0 (1.04)	25.29	1, 156	< .001	0.14

Note: Statistically significant p -values ($p < .05$) are shown in bold text.

after initial failure compared to after initial success (Table 4; See also Figure 2).

Participants reported a non-significant decrease (<4) in self-efficacy and in perceived usefulness after initial failure ($ps > .149$), and they reported an increase (>4) in self-efficacy and perceived usefulness after initial success ($ps < .001$). Indeed, the distributions in Figure 2 shows great variation in participants' relative self-efficacy and perceived usefulness after the initial failure (orange points): some reported an increase (<4) and some a decrease (>4). In contrast, after initial success (blue points), the majority of participants reported an increase (>4) in self-efficacy and perceived usefulness. Relative self-efficacy and perceived usefulness were lower after failure compared to after success (Table 4).

Mediation analyses

See Table 5 for the results of the mediation analysis. Relative intentions, self-efficacy, and perceived usefulness all significantly mediated the effect of regulation on failure T2 and success T2, but the indirect effects for intention were markedly smaller compared to self-efficacy and perceived usefulness.

TABLE 5 Study 2: regression coefficients and 95% bias-corrected bootstrapped confidence intervals (CI) for the indirect effects of regulation on failure T2 and success through self-efficacy, motivation, and/or perceived usefulness

Mediator	Regulation–Failure T2		Regulation–Success T2	
	B	95% CI	B	95% CI
Intention	0.55	[0.02, 2.05]	−0.11	[−0.31, −0.01]
Self-efficacy	0.83	[0.08, 2.19]	−0.35	[−0.66, −0.14]
Perceived usefulness	0.83	[0.11, 2.11]	−0.34	[−0.60, −0.14]

Note: Statistically significant confidence intervals (excluding zero) are shown in bold text.

4.3 | Discussion

Results of Study 2 confirmed our hypothesis that experiencing failure may indeed trigger subsequent failure, and that relative self-efficacy after the initial failure mediated this effect: lower relative self-efficacy after the initial failure was associated with more subsequent failure.

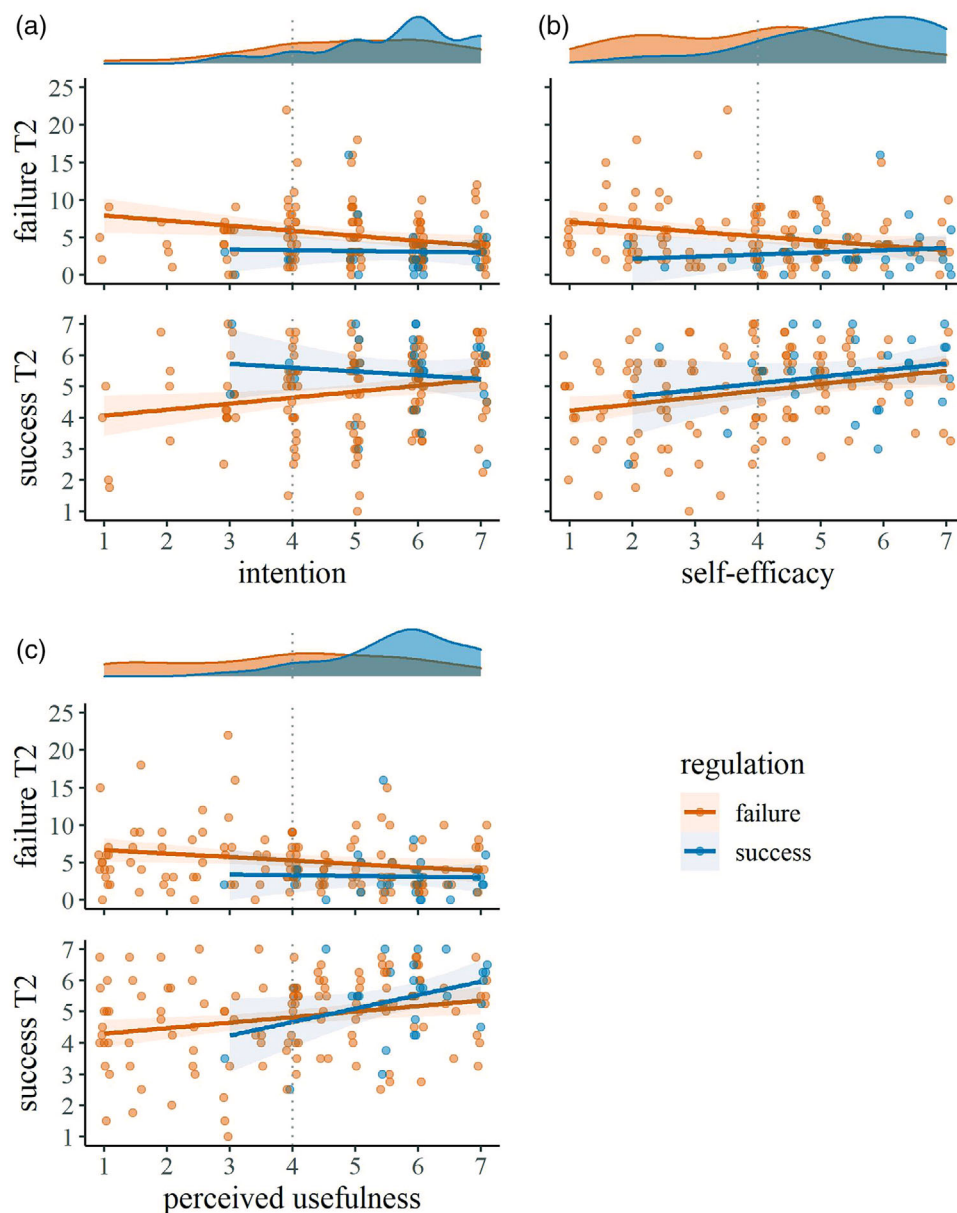


FIGURE 2 Study 2: density distributions of intention (a), self-efficacy (b), and perceived usefulness (c) and their relationship to failure T2 and success T2, separated for regulation conditions. Light blue and light orange areas reflect 95% confidence intervals. The grey dotted line shows the midpoint (4) of the scale. *Note.* For clarity reasons, the scatterplots do not show one extreme score of 39 on failure T2. However, this score is taken into account in the fitting of the regression line.

Even though on average, participants' self-efficacy did not significantly decrease (<4), Figure 2 demonstrates that many people experienced a decrease in self-efficacy after they experienced failure, and that these are also more likely to experience more subsequent failure and lower success. Moreover, people who were able to maintain (~ 4) or even increase (> 4) their self-efficacy after the initial failure showed more or less similar levels of failure and success in the following days compared to people who had experienced success at T1 (Figure 2). These findings align with recent literature suggesting that interventions to boost self-efficacy upon experiencing failure are helpful in increasing subsequent self-regulation performance and avoiding a setback or relapse effect (Roordink et al., 2021).

In addition, in Study 2 we also found evidence for a mediating effect of perceived usefulness: lower relative perceived usefulness after the initial failure was associated with more subsequent failure. In addition, although participants unexpectedly reported an increase in intention to self-regulate after they experienced failure, people who experienced stronger relative intentions were more successful at self-regulation at T2. However, replication of these findings, in particular of the findings regarding perceived usefulness which were not listed as confirmatory and do not align with the literature on the disinhibition effect (Herman & Mack, 1975), is warranted before drawing any conclusions. Such a replication should also address the unequal division of participants across the failure and success condition.

5 | STUDY 3

In Study 3 we again tested the role of intentions, self-efficacy, and perceived usefulness, but this time in the domains of exercising, saving, and eating. Moreover, to ensure more equal groups for investigating the overarching main effect of failure versus success, we compared participants who failed to act in line with their goal in the past 24 h to participants who were successful in adhering to their goal in this same time period. To diminish retrospective bias, we incorporated daily assessments of self-regulation failure and success as dependent measures. The study was preregistered at the open science framework (OSF): https://osf.io/zqhxw/?view_only=125075a71ce441728fdeb472d21ea51d.

5.1 | Methods

5.1.1 | Participants

We recruited through Prolific a total of 1140 native English speaking participants, aged 18–65, currently residing in the United Kingdom who had not participated in our other studies on the setback effect. Three subsamples (i.e., eating subsample, exercise subsample, and saving subsample) were created based on participants' responses to three goal questions: 'Do you currently in your everyday life (at least for the past 7 days) actively pursue the following goal: [restrict your unhealthy food intake in any way/increase or maintain a certain level of exercising in any way/restrict spending money on certain products of activities in any way]?' Participants were included in *all* subsamples for which they responded 'yes' to the goal question. Participants responding 'no' to all goal questions were excluded.

The sample size was based on a-priori power analyses for all confirmatory hypotheses for the three goal domains separately in G*Power, the largest required sample size being $N = 274$ for the effect of regulation on the outcome measures and the mediator variables, aiming for a power of 0.80 ($\alpha = 0.05$) and a minimally meaningful effect size of $f = 0.17$ (small to medium). Anticipating a subsample-size of at least 60% and accounting for a drop-out of 60%,⁶ we needed to recruit a sample of 1140 to maintain the final subsample of $N = 274$. Actual drop-out was 41%.

We had an eating subsample of $N = 415$ (301 females) with an average age of 38.06 years ($SD = 12.35$) and an average BMI of 26.60 ($SD = 6.25$). We had an exercise subsample of $N = 366$ (236 females) with an average age of 37.59 years ($SD = 11.72$) and an average BMI of 26.06 ($SD = 5.81$). We had a saving subsample of $N = 525$ (371 females) with an average age of 36.83 years ($SD = 11.69$) and an average BMI of 26.15 ($SD = 6.04$). Participants who dropped out were significantly younger ($\Delta M = 2.92$, $p < .001$), reported significantly lower (relative)

self-efficacy after exercise failure ($\Delta M = 0.22$, $p = .038$), and were more likely to be in the failure condition in the eating domain ($p = .037$) compared to participants who completed the study.

5.1.2 | Design and procedure

Study 3 followed a similar procedure as Study 2. For each relevant goal, participants were asked whether they had failed in the past 24 h. Based on pilot data, this window of 24 h was expected to yield two approximately equal groups responding 'yes' (failure condition) and 'no' (success condition). Next, participants recalled the most recent situation in which they violated the goal (failure condition) or had resisted violating the goal (success condition). Then, they answered questions on intention, self-efficacy, and perceived usefulness relative to before their most recent failure or success (framed as 'after you violated your [goal]' for the failure condition, and as 'after you successfully adhered to your [goal]' in the success condition).

Participants who indicated that they adhered to (a subset of) the three goals in part 1 were invited to participate in part 2 (T2). In contrast to Study 2, part 2 involved four *daily* reports on failure and success (Day 2 to Day 5 of the procedure). On each day, at 7 p.m. local time, participants were invited to complete the survey for that day within 4 h. Participants indicated the day and date, visualized this on the provided picture of a calendar, and indicated at what time they filled in yesterday's questionnaire to facilitate recall. Next, they were asked to consider the timeframe between filling in the previous questionnaire and the current questionnaire, and to indicate their failure and success regarding the respective goal(s) for this timeframe. Participants received £3.50.

5.1.3 | Materials

Materials were similar to Study 2. For the exercising domain and the saving domain, all materials were adapted to the respective behaviours. In addition, for all domains, the following adaptations were made.

Regulation

To assess regulation (failure or success) at T1, participants were asked whether they failed the respective goal 'in the past 24 h'.

Self-efficacy, perceived usefulness and intention

Intention was assessed with two items instead of one item (for details, see supplementary files; $r_{\text{eating}} = .70$, $r_{\text{exercising}} = .69$, $r_{\text{saving}} = .64$). The questions on self-efficacy were slightly adapted ($r_{\text{eating}} = .73$, $r_{\text{exercising}} = .75$, $r_{\text{saving}} = .65$): one of the two items was replaced to tap more specifically into the concept self-efficacy (for details, see supplementary files). For perceived usefulness, the same two items were used as in Study 2 ($r_{\text{eating}} = .73$, $r_{\text{exercising}} = .71$, $r_{\text{saving}} = .64$).

Failure and success at T2

On each day of part 2, participants were asked to think about the time between filling in yesterday's questionnaire and now, and to answer

⁶ The estimated subsample size was based on the aforementioned pilot study of $N = 100$. The expected dropout-rate was based on an examination of the first 100 participants of the final sample. Please note that this examination was planned and executed in order to determine drop-out rate and the variance in the outcome measures. No other statistical tests were performed on this subsample.

the items for failure T2 and success T2. *Failure T2* was calculated as the sum of four daily failure measures, using the question 'How many times during this time did you [eat something that was not in line with your specific dieting goal/skip scheduled exercise sessions that was in your specific exercise schedule/spend money on a product or activity that was not in line with your specific goal to restrict spending money]?'. *Success T2* was calculated as the mean of four daily measures, using the question 'To what extent do you feel like you successfully adhered to your [dieting goal/ exercise goal/ goal to restrict spending money] during this time?'

5.1.4 | Data-analysis

Within each domain we conducted the same analyses as in Study 2. In addition, we conducted exploratory (non-preregistered) analyses to get preliminary insights into whether our hypotheses can also be confirmed on the within-participant level, across behavioural domains, as this would provide additional insights to rule out general (i.e., across domains) levels of intention, self-efficacy, and perceived usefulness as third variable explanation for the association between regulation and failure and success T2.⁷ We conducted these within-participants analyses on a subsample of participants who had data on two ($N = 278$) or three ($N = 184$) behavioural domains.

First, we fitted three separate mixed-effects regression models predicting intention, self-efficacy, or perceived usefulness with regulation (failure vs. success), domain (eating vs. exercising vs. saving), and a random intercept for participant as predictors. Next, we fitted six separate mixed effects regression models predicting failure T2 or success T2, with intentions/self-efficacy/perceived usefulness, domain (eating vs. exercising vs. saving) and a random intercept for participants as predictors.

5.2 | Results

See Table 6 for descriptives and intercorrelations. Statistics are presented in Tables 7 and 8. Within each domain, we performed a square root transformation on failure T2 to normalize the distribution. Means and standard deviations are presented for the untransformed variables.

5.2.1 | Eating domain

Descriptives and intercorrelations

In total, 304 participants were in the failure condition and 111 participants were in the success condition. Due to technical issues several participants had missing responses, leaving $N = 414$ for the analyses on failure T2, and $N = 407$ for the analyses on self-efficacy, intention, and perceived usefulness. Age and BMI did not differ between condi-

tions ($ps \geq 0.429$). Intention was significantly and negatively related to failure T2 ($p = .022$), but not to success T2 ($p = .315$). Self-efficacy and perceived usefulness were significantly and negatively related to failure T2 ($ps \leq 0.003$) and positively related to success T2 ($ps < .001$).

Main analyses

Participants reported significantly more failure T2 and lower success T2 after initial failure than after success (Table 7). Participants reported an increase (>4) in intention both after initial failure ($p < .001$) and after initial success ($p < .001$). Relative intention did not differ between the failure and success condition (Table 7).

Participants reported a non-significant decrease (<4) in self-efficacy after initial failure ($p = .317$) and an increase (>4) after initial success ($p < .001$). Relative self-efficacy was lower after initial failure compared to after success (Table 7). Participants reported an increase (>4) in perceived usefulness both after initial failure ($p < .001$) and after initial success ($p < .001$). Relative perceived usefulness was lower after failure compared to after success (Table 7).

The mediation analyses (Table 8) indicated that self-efficacy significantly mediated the effect of regulation on both outcome measures, and perceived usefulness significantly mediated the effect of regulation on failure T2, but not on success T2.

5.2.2 | Exercise domain

Descriptives and intercorrelations

In total, 83 participants were in the failure condition and 283 participants were in the success condition. Due to technical issues several participants had missing responses, leaving $N = 364$ for the analyses on failure T2, and $N = 363$ for the analyses on self-efficacy, intention and perceived usefulness. Independent t -tests indicated that age and BMI did not differ between conditions ($ps \geq .078$). Self-efficacy, perceived usefulness and intention were significantly and positively related to success T2 ($ps \leq .001$), but not to failure T2 ($ps \geq .132$).

Main analyses

Participants reported significantly more failure T2 and lower success T2 after initial failure than after success (Table 7). Participants reported an increase (>4) in intention both after initial failure ($p < .001$) and after initial success ($p < .001$). Relative intentions were lower after failure compared to after success (Table 7).

Participants reported non-significant increase (>4) in self-efficacy and a non-significant decrease (<4) in perceived usefulness after initial failure ($ps > .191$), and a significant increase (>4) in both after initial success ($ps < .001$). Relative self-efficacy and perceived usefulness were lower after initial failure compared to after initial success (Table 7).

The mediation analyses (Table 8) indicated that relative intentions, self-efficacy, and perceived usefulness significantly mediated the effect of regulation on success T2, but not in failure T2. The indirect effect through intentions was markedly smaller compared to self-efficacy and perceived usefulness.

⁷ We thank an anonymous reviewer for suggesting this additional exploratory analysis.

TABLE 6 Study 3: means (M), standard deviations (SD) and intercorrelations

	1	2	3	4	5	6	7
Eating domain							
Age (1)							
BMI (2)	.10						
Intention (3)	.16	.04					
Self-efficacy (4)	0.07	.03	.28				
Perceived usefulness (5)	0.12	0.01	0.43	.47			
Failure T2 (6)	−.02	0.09	−.11	−0.29	−0.15		
Success T2 (7)	.00	−0.15	0.05	.31	0.18	−0.68	
M	38.06	26.60	5.06	4.30	4.92	4.39	4.79
SD	12.35	6.25	1.33	1.49	1.55	4.60	1.39
Exercising domain							
Age (1)							
BMI (2)	.11						
Intention (3)	0.09	0.01					
Self-efficacy (4)	.11	.03	.56				
Perceived usefulness (5)	0.10	.07	.45	0.61			
Failure T2 (6)	−0.01	−.01	−.03	−0.08	−0.08		
Success T2 (7)	0.01	−0.04	0.18	0.21	0.17	−0.66	
M	37.59	26.06	5.33	5.14	5.22	1.05	5.47
SD	11.72	5.81	1.17	1.30	1.37	1.32	1.35
Saving domain							
Age (1)							
BMI (2)	.17						
Intention (3)	−.05	0.05					
Self-efficacy (4)	.07	−0.02	0.30				
Perceived usefulness (5)	−.05	−0.03	.44	0.33			
Failure T2 (6)	−0.13	−.05	−0.08	−0.16	−0.12		
Success T2 (7)	.08	−0.01	0.03	0.14	0.09	−0.60	
M	36.83	26.15	4.82	5.26	5.54	1.91	5.79
SD	11.69	6.04	1.33	1.17	1.20	2.70	1.08

Note: Correlation coefficients in bold are significant ($p < .05$)

5.2.3 | Saving domain

Descriptives and intercorrelations

In total, 218 participants were in the failure condition and 307 participants were in the success condition. Due to technical issues several participants had missing responses, leaving $N = 523$ for the analyses on self-efficacy, intention, and perceived usefulness. Independent t -tests indicated that age and BMI were not different between conditions ($ps \geq .759$). Self-efficacy was significantly and negatively related to failure T2 ($p < .001$) and positively related to success T2 ($p = .001$). Intention was not significantly related to any of the outcome measures ($ps \geq .061$). Perceived usefulness was significantly

and negatively related to failure T2 ($p = .006$), but not to success T2 ($p = .052$).

Main analyses

Participants reported significantly more failure T2 and lower success T2 after initial failure than after success (Table 7). Participants reported a (small) increase (>4) in intentions, self-efficacy, and perceived usefulness both after initial failure ($ps < .034$), and after initial success ($p < .001$). Relative intention did not differ between the regulation conditions (Table 7). Relative self-efficacy and perceived usefulness were lower after initial failure compared to after initial success (Table 7).

TABLE 7 Study 3: means (*M*), standard deviations (*SD*), and statistics of the MANOVAs and follow-up ANOVAs on the effect of regulation on failure T2 and success T2, and on intention, self-efficacy, and perceived usefulness

	Regulation T1		(M)ANOVA results			
	Failure group	Success group	<i>F</i>	<i>df</i> 1, <i>df</i> 2	<i>p</i>	partial η^2
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)				
Eating domain	<i>N</i> = 304	<i>N</i> = 111				
Multivariate			22.00	2, 411	<.001	
Failure T2	4.99 (4.87)	2.77 (3.28)	37.85	1, 412	<.001	0.08
Success T2	4.54 (1.41)	5.48 (1.10)	40.44	1, 413	<.001	0.09
Multivariate			32.44	3, 406	<.001	
Intention	5.01 (1.34)	5.20 (1.31)	1.72	1, 408	.190	0.004
Self-efficacy	3.92 (1.41)	5.34 (1.19)	88.73	1, 408	<.001	0.18
Perceived usefulness	4.67 (1.58)	5.56 (1.24)	32.18	1, 408	<.001	0.07
Exercising domain	<i>N</i> = 83	<i>N</i> = 283				
Multivariate			6.39	2, 361	.005	
Failure T2	1.39 (1.39)	0.95 (1.29)	10.66	1, 362	.001	0.03
Success T2	5.13 (1.40)	5.57 (1.32)	6.84	1, 364	.009	0.02
Multivariate			49.89	3, 359	<.001	
Intention	4.98 (1.35)	5.43 (1.10)	9.95	1, 361	.002	0.03
Self-efficacy	4.20 (1.38)	5.41 (1.14)	64.20	1, 361	<.001	0.15
Perceived usefulness	3.90 (1.45)	5.61 (1.08)	133.39	1, 361	<.001	0.27
Saving domain	<i>N</i> = 218	<i>N</i> = 307				
Multivariate			19.71	2, 522	<.001	
Failure T2	2.62 (3.55)	1.40 (1.72)	38.82	1, 523	<.001	0.07
Success T2	5.51 (1.17)	5.99 (0.97)	26.38	1, 523	<.001	0.05
Multivariate			41.19	3, 519	<.001	
Intention	5.31 (1.10)	5.22 (1.15)	0.77	1, 521	.381	< .01
Self-efficacy	4.19 (1.32)	5.27 (1.14)	99.79	1, 521	<.001	0.16
Perceived usefulness	5.35 (1.34)	5.67 (1.06)	8.88	1, 521	.003	0.02

Note: Statistically significant *p*-values (*p* < .05) are shown in bold text.

TABLE 8 Study 3: regression coefficients and 95% bias-corrected bootstrapped confidence intervals (CI) for the indirect effects of regulation on failure T2 and success through self-efficacy, intention, and/or perceived usefulness

Mediator	Regulation–Failure T2		Regulation–Success T2	
	<i>B</i>	95% CI	<i>B</i>	95% CI
Eating domain				
Self-efficacy	1.11	[0.59, 1.77]	−0.30	[−0.48, −0.16]
Perceived usefulness	0.28	[0.02, 0.61]	−0.08	[−0.19, 0.01]
Exercising domain				
Intention	0.01	[−0.04, 0.07]	−0.08	[−0.20, −0.02]
Self-efficacy	0.03	[−0.13, 0.17]	−0.23	[−0.39, −0.06]
Perceived usefulness	0.01	[−0.20, 0.20]	−0.23	[−0.45, −0.05]
Saving domain				
Self-efficacy	0.19	[−0.07, 0.63]	−0.06	[−0.14, 0.02]
Perceived usefulness	0.07	[−0.02, 0.35]	−0.02	[−0.06, 0.01]

Note: Statistically significant confidence intervals (excluding zero) are displayed in bold text.

The mediation analyses (Table 8) indicated that relative self-efficacy and perceived usefulness did not significantly mediate the effect of regulation on the outcomes.

5.2.4 | Exploratory secondary analyses: Within-participant effects

A full overview of statistics is presented in Table S1 and S2. In summary, the within-participant effect of regulation on relative intentions was not significant, 95% CI $[-0.25, 0.06]$. The within-participant effects of regulation on relative self-efficacy and on perceived usefulness were significant, with lower relative self-efficacy (CI $[-1.36, -1.03]$) and lower relative perceived usefulness (CI $[-0.99, -0.65]$) after initial failure compared to after initial success.

The within-participant effect of intention on failure T2 (95% CI $[-0.07, 0.01]$) and success T2 (CI $[-0.00, 0.12]$) were not significant. The within-participant effects of self-efficacy on failure T2 and success T2 were significant, with higher relative self-efficacy being associated with less failure T2 (CI $[-0.16, -0.09]$) and more success T2 (CI $[0.13, 0.24]$). The within-participant effect of perceived usefulness on failure T2 and success T2 were also significant, with higher relative perceived usefulness being associated with less failure T2 (CI $[-0.12, -0.03]$) and more success T2 (CI $[0.07, 0.18]$).

5.3 | Discussion

Results from Study 3 confirmed findings from Study 2 that a setback increases the chance of subsequent failure, also for exercising and saving goals. In addition, results confirmed the roles of self-efficacy and perceived usefulness as underlying processes in this setback effect in eating behaviour. This suggests that interventions to maintain or boost self-efficacy upon experiencing failure could be effective in protecting subsequent self-regulation performance. Similar to Study 2, the results of Study 3 also suggest a mediating role for perceived usefulness to self-regulate, but again participants generally reported an increase in perceived usefulness after they experienced failure. This suggests mainly a positive effect of success through an increase in perceived usefulness, rather than a setback effect fuelled by a what-the-hell effect.

Results for exercising showed a similar picture, although the mediating roles of self-efficacy and perceived usefulness were only confirmed for subsequent success, and participants generally reported an increase or no change in these cognitions after experiencing a setback. In the saving domain none of the mediators were significant. A plausible explanation for these weaker/non-significant effects could be the limited variance in the frequency of failure in these domains.

Interestingly, additional exploratory analyses of the within-participant effects confirmed that across domains, when a participant experienced a setback they experienced lower relative self-efficacy and perceived usefulness than when they experienced success. In addition,

lower-than-usual relative self-efficacy and perceived usefulness increased the likelihood of subsequent failure across domains. These within-participant findings disprove the alternative explanation that domain-general levels of self-efficacy and perceived usefulness explain both failure and success on T1 and T2.

6 | GENERAL DISCUSSION

Although the potential of a single setback in self-regulation snowballing into full-blown self-regulation failure has been proposed as a key threat to long-term self-regulation success (Baumeister & Heatherton, 1996; Wagner & Heatherton, 2015), the empirical evidence for such a setback effect is mixed and limited to pathological behaviour and/or specific contexts. Moreover, it is unclear what is driving this effect. In the present article, we investigated (a) whether the experience of a setback indeed increases the chance of subsequent failure in the context of everyday self-regulation, and (b) what psychological mechanisms underly this detrimental effect. Results across all studies and behavioural domains, that is, eating, exercising, and saving, demonstrated that experiencing a single setback in self-regulation—compared to success—increases the likelihood of subsequent failure and decreases the chances at subsequent self-regulation success. Our results most consistently point to changes in people's self-efficacy after the initial failure or success as an underlying mechanism of this effect. We also found a potential role for people's perceptions of the usefulness to self-regulate, and to a lesser extent people's intentions to self-regulate, but these results were less consistent.

Our findings build on Wenzel et al. (2020) in going beyond the classic belief that resisting temptations impairs subsequent self-control (i.e., the classic ego-depletion effect). In contrast, our findings suggest that successfully resisting a temptation is beneficial for future self-regulation success, and experiencing a single setback irrationally makes people more prone to failure in subsequent situations. These insights align with concerns that the setback effect may be a relevant threat to everyday self-regulation (Wagner & Heatherton, 2015) and suggest that the scientific field needs to shift the focus from preventing self-regulation failure, to understanding what happens in the aftermath of a single setback.

Specifically, the field should acknowledge that experiencing some setbacks is inevitable when pursuing long-term goals. Indeed, our findings suggest that instances of self-regulation failure are highly common. While we aimed for an equal division in regulation conditions in Study 2, only 19% of the respondents had been successful in adhering to their diets in the preceding 48 h, and 81% of females reported to have experienced a setback in the previous 48 h. Similarly, in Study 3, which targeted the past 24 h, 73% of participants pursuing an eating goal, 23% of participants pursuing an exercising goal, and 42% of participants pursuing a saving goal reported to have failed on their respective goal within this timeframe. So, setbacks are indeed highly common and inevitably experiencing a setback impairs subsequent self-regulation

success. These observations emphasize the importance of investigating people's responses to failure, to ultimately develop interventions that help people prevent an irrational and maladaptive downward spiral of failure.

To identify potential intervention targets, we explored underlying mechanisms of the setback effect. Our results most consistently suggest that a single setback increases the chance of subsequent failure due to a decrease in self-efficacy after the setback. After participants experienced a setback in regulating their eating behaviour, many felt less confident that they were able to regulate their eating behaviour, and consequently they were more prone to eat or do something that was not in line with their goal in subsequent situations. These findings contradict the common idea that people have a need to behave consistent with their beliefs and, and therefore, they attribute failure to external factors, whereas they attribute successes to themselves (i.e., self-serving bias; Mezulis et al., 2004). In line with the abstinence violation effect (Marlatt & Witkiewitz, 2005), participants in our study were likely to attribute the setback to internal, stable reasons, which lowered their feelings of personal control. Moreover, the finding that people experience a decrease in self-efficacy after a setback, and an increase in self-efficacy after success, is consistent with the idea that people derive their self-efficacy from previous performance (Bandura, 1978).

Interestingly, on average, self-efficacy did not always decrease (<4) after the initial failure: even though many people experienced a decrease in self-efficacy after the initial failure, many people were also able to maintain or even boost their self-efficacy after failure (see Figures 1 and 2). These people seemed to experience similar levels of failure and success in the subsequent days compared to people who were initially successful. These findings support the idea that interventions aiming to prevent the setback effect should aim to help people maintain or boost their self-efficacy after they experience a setback. Moreover, these insights add to the research basis on self-efficacy and feelings of personal control as crucial precursors of relapse into physical activity, unhealthy eating (Roordink et al., 2021), and addiction (Marlatt & Witkiewitz, 2005).

In addition to self-efficacy, we also found that the setback effect is partly fuelled by changes in people's perceptions of the usefulness to self-regulate and, in some instances, changes in their intentions to self-regulate. Yet, on average, individuals generally experienced an increase in intentions and perceived usefulness to self-regulate compared to before the setback. In other words, after experiencing a setback, individuals on average *wanted* to self-regulate even more, and perceived it to be *more useful* to self-regulate in future situations.

This contradicts the classic idea of a what-the-hell effect (Herman & Mack, 1975), suggesting that after a diet violation people perceive future self-regulation to have become futile. It also contradicts the cognitive dissonance effect (Stone & Cooper, 2001), suggesting that people may attribute a setback to a lack of motivation to self-regulate. Interestingly, it seems that people mainly experience a strong increase in intentions and perceived usefulness to self-regulate after they are successful, which in turn increases the chances of subsequent success, rather than decreased intentions and what-the-hell cognitions

fuelling a setback effect. These findings suggest that intentions and the perceived usefulness to self-regulate are less promising intervention targets, as there is little room left for change. Moreover, intentions to self-regulate failed to consistently predict actual subsequent self-regulation failure, and the effect was smaller than for self-efficacy and perceived usefulness.

As such, based on our findings, the most promising intervention target for preventing a setback effect would be people's self-efficacy. Specifically, when experiencing an inevitable setback, individuals need an intervention strategy to maintain or increase their confidence in their abilities to regulate their behaviour in upcoming self-regulation dilemmas. An example of such an intervention could be to help people refocus on previous successes (e.g., 'Even though I just failed, yesterday I resisted three temptations, so I know that I am capable of adhering to my diet'), a strategy that has been successful in enhancing self-efficacy for physical activity (Ashford et al., 2010). Interventions could also more indirectly target self-efficacy by guiding people's attributions regarding the cause of the setback (Seligman et al., 1979). Adriaanse and ten Broeke (2022) designed and tested a brief mindset intervention based on this idea. Participants were informed that setbacks may occur and instructed to attribute future incidents of failure to external causes. Results demonstrated that this brief intervention led to a reduction in subsequent self-regulation failure and there was tentative evidence that this effect was partly mediated by protecting people's self-efficacy upon experiencing a setback.

7 | LIMITATIONS AND AVENUES FOR FUTURE RESEARCH

Several limitations should be noted. First, the studies in the present article mainly served as an initial exploration of an understudied topic. As a result a relative high number of statistical tests were performed, which could have enhanced the chance at false positives. It is evident that one should be cautious with drawing firm conclusions without additional future replications.

Secondly, as we did not experimentally manipulate failure in our studies we cannot rule out the possibility that a third variable is responsible for initial failure versus success as well as the degree of failure (or success) in the follow-up measure. However, baseline scores on BMI and restrained eating did not differ between the success and failure group in Study 2 and Study 3. Moreover, we ruled out the possibility that (general) intentions, self-efficacy, and perceived usefulness at baseline alternatively explained both failure at T1 and at T2 in two ways: (a) We used a 'compared to before' question format to specifically assess *changes* in these cognitions (i.e., inherently controlling for baseline intentions, self-efficacy and perceived usefulness) (b) We found preliminary evidence for the associations between failure and changes in self-efficacy and perceived usefulness, and between self-efficacy and perceived usefulness and subsequent failure, on the within-participant level.

Yet, in order to rule out the third variable explanation, future research could investigate the setback effect on a within-participant,

within-domain level, using a study design that better allows researchers to examine the direction of the effect, such as Ecological Momentary Assessment (e.g., Hofmann et al., 2012). Future studies should also develop research paradigms that do enable the manipulation of failure versus success in real behaviour, while carefully considering potential unwanted side effects, such as artificial settings, or (unintendedly) providing participants with a situational explanation. For instance, previous studies used false feedback to induce feelings of goal-violation (e.g., Gantman et al., 2017).

Finally, given the exploratory nature of the present studies, and the way in which mechanisms and outcome measures were assessed, the current results did not allow for detailed investigations into how intention, self-efficacy and perceived usefulness might be (causally) interdependent in fuelling the setback effect. In addition, we did not examine the potential moderating role of causal attributions (Seligman et al., 1979): according to the abstinence violation effect, a setback effect is most likely to occur if people attribute the initial setback to internal and uncontrollable causes (Curry et al., 1987). Altogether, these comprise interesting questions for future research.

8 | CONCLUDING REMARKS

As ironically endorsed by our failed attempt to collect an equal amount of success and failure responses in Study 2, experiencing failure is simply inevitable. The present research confirmed that although setback in self-regulation may in itself be harmless, depending on subsequent changes in relevant cognitions, such as people's confidence in their abilities to regulate their behaviour, a single setback could be the starting point of a slippery slope towards more failure. We found that this is not necessarily because people no longer want to self-regulate, or perceive it to be useless, but because they feel less capable to regulate their behaviour in subsequent situations when they experienced a setback. As such, psychologists should provide people with intervention strategies to apply when they experience an inevitable setback, to help them maintain their confidence and prevent a downward spiral of self-regulation failure. More broadly, with the present research, we aimed to initiate a shift in scientific focus from preventing self-regulation failure, to investigating people's responses to inevitable setbacks.

CONFLICT OF INTEREST

The authors have no known conflict of interest to disclose.

DATA AVAILABILITY STATEMENT

All data that we used for our analyses, and R code for data processing, analysis and visualization, will be made available on the Open Science Framework: https://osf.io/58dfj/?view_only=492e222dcf8941bf95dd20b25903b400.

ETHICS APPROVAL STATEMENT

Studies reported in this article were approved by the Ethics Committee of the Faculty of Social and Behavioural Sciences of Utrecht University.

REFERENCES

- Adriaanse, M. A., & ten Broeke, P. (2022). Beyond prevention: Regulating responses to self-regulation failure to avoid a set-back effect. *Applied Psychology: Health and Well-Being*, 14(1), 278–293. <https://doi.org/10.1111/aphw.12302>
- Adriaanse, M. A., Oettingen, G., Gollwitzer, P. M., Hennes, E. P., De Ridder, D. T. D., & De Wit, J. B. F. (2010). When planning is not enough: Fighting unhealthy snacking habits by mental contrasting with implementation intentions (MCII). *European Journal of Social Psychology*, 40(7), 1277–1293. <https://doi.org/10.1002/ejsp.730>
- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the theory of planned behavior. *Journal of Applied Social Psychology*, 32(4), 665–683. <https://doi.org/10.1111/j.1559-1816.2002.tb00236.x>
- Ashford, S., Edmunds, J., & French, D. P. (2010). What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *British Journal of Health Psychology*, 15(2), 265–288. <https://doi.org/10.1348/135910709x461752>
- Bandura, A. (1978). Self-efficacy: Toward a unifying theory of behavioral change. *Advances in Behaviour Research and Therapy*, 1(4), 139–161. [https://doi.org/10.1016/0146-6402\(78\)90002-4](https://doi.org/10.1016/0146-6402(78)90002-4)
- 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(5), 1252–1265. <https://doi.org/10.1037/0022-3514.74.5.1252>
- Baumeister, R. F., & Heatherton, T. F. (1996). Self-regulation failure: An overview. *Psychological Inquiry*, 7(1), 1–15. https://doi.org/10.1207/s15327965pli0701_1
- Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139174794>
- Curry, S., Marlatt, G. A., & Gordon, J. R. (1987). Abstinence violation effect: Validation of an attributional construct with smoking cessation. *Journal of Consulting and Clinical Psychology*, 55(2), 145–149. <https://doi.org/10.1037/0022-006X.55.2.145>
- de Ridder, D., Adriaanse, M., & Fujita, K. (2017). *Routledge international handbook of self-control in health and well-being*. Routledge.
- Eisenhower, D., Mathiowetz, N. A., & Morganstein, D. (1991). Recall error: Sources and bias reduction techniques. In R. Biemer, R. Groves, L. Lyberg, N. Mathiowetz, & Sudman (Eds.), *Measurement errors in surveys*. Wiley 125–144.
- Faul, F., Erdfelder, E., Lange, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE.
- Gantman, A. P., Adriaanse, M. A., Gollwitzer, P. M., & Oettingen, G. (2017). Why did I do that? Explaining actions activated outside of awareness. *Psychonomic Bulletin & Review*, 24(5), 1563–1572. <https://doi.org/10.3758/s13423-017-1260-5>
- Herman, C. P., & Mack, D. (1975). Restrained and unrestrained eating. *Journal of Personality*, 43(4), 647–660. <https://doi.org/10.1111/1467-6494.ep8970396>
- Herman, C. P., Polivy, J., & Esses, V. M. (1987). The illusion of counter-regulation. *Appetite*, 9(3), 161–169. [https://doi.org/10.1016/S0195-6663\(87\)80010-7](https://doi.org/10.1016/S0195-6663(87)80010-7)
- Herman, C. P., & Polivy, J. (1980). Restrained eating. In A. J. Stunkard (Ed.), *Obesity* (pp. 108–225). Saunders.
- Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology*, 102(6), 1318–1335. <https://doi.org/10.1037/a0026545>
- Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating models of self-regulation. *Annual Review of Psychology*, 72(1), 319–345. <https://doi.org/10.1146/annurev-psych-061020-105721>
- Jansen, A., Merckelbach, H., Oosterlaan, J., Tuiten, A., & Van den Hout, M. (1988). Cognitions and self-talk during food intake of restrained and

- unrestrained eaters. *Behaviour Research and Therapy*, 26(5), 393–398. [https://doi.org/10.1016/0005-7967\(88\)90072-1](https://doi.org/10.1016/0005-7967(88)90072-1)
- Jansen, A., Nederkoorn, C., Van Baak, L., Keirse, C., Guerrieri, R., & Havermans, R. (2009). High-restrained eaters only overeat when they are also impulsive. *Behaviour Research and Therapy*, 47(2), 105–110. <https://doi.org/10.1016/j.brat.2008.10.016>
- Kabacoff, R. L. (2011). *R in action: Data analysis and graphics with R*. Manning Publications Co.
- Marlatt, G. A., & Witkiewitz, K. (2005). Relapse prevention for alcohol and drug problems. In G. A. Marlatt, & D. M. Donovan (Eds.), *Relapse prevention: maintenance strategies in the treatment of addictive behaviors* (pp. 1–44). Guilford Press.
- Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, 130(5), 711–747. <https://doi.org/10.1037/0033-2909.130.5.711>
- Michie, S., Atkins, L., & West, R. (2014). The behaviour change wheel: A guide to designing interventions. Silverback.
- Milyavskaya, M., & Inzlicht, M. (2017). Attentional and motivational mechanisms of self-control. *The routledge international handbook of self-control in health and well-being*. Routledge.
- Roordink, E. M., Steenhuis, I. H. M., Kroeze, W., Schoonmade, L. J., Sniehotta, F. F., & van Stralen, M. M. (2021). Predictors of lapse and relapse in physical activity and dietary behaviour: A systematic search and review on prospective studies. *Psychology & Health*, 0(0), 1–24. <https://doi.org/10.1080/08870446.2021.1981900>
- Seligman, M. E., Abramson, L. Y., Semmel, A., & von Baeyer, C. (1979). Depressive attributional style. *Journal of Abnormal Psychology*, 88, 242–247. <https://doi.org/10.1037/0021-843X.88.3.242>
- Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology*, 37(3), 228–243. <https://doi.org/10.1006/jesp.2000.1446>
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics*. Pearson Education.
- Tomiya, A. J., Moskovich, A., Haltom, K. B., Ju, T., & Mann, T. (2009). Consumption after a diet violation: Disinhibition or compensation? *Psychological Science*, 20(10), 1275–1281. <https://doi.org/10.1111/j.1467-9280.2009.02436.x>
- Wagner, D. D., & Heatherton, T. F. (2015). Self-regulation and its failure: The seven deadly threats to self-regulation. In M. Mikulincer & P. R. Shaver (Eds.), *APA handbook of personality and social psychology: vol 1. Attitudes and social cognition*. American Psychological Association 805–842.
- Wenzel, M., Rowland, Z., Hofmann, W., & Kubiak, T. (2020). Setbacks in self-control: Failing not mere resisting impairs subsequent self-control. *Social Psychological and Personality Science*, 11(6), 782–790. <https://doi.org/10.1177/1948550619888875>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: ten Broeke, P., & Adriaanse, M. A. (2022). Understanding the setback effect in everyday self-regulation. *European Journal of Social Psychology*, 1–17. <https://doi.org/10.1002/ejsp.2931>

APPENDIX

Data Handling Protocol

Outliers

For each analysis, data was screened on univariate outliers and multivariate outliers. Univariate outliers were defined as (1) having an absolute z-score larger than 3.29 and (2) being visually disconnected from the distribution in the histogram. For group-wise analyses (e.g., ANOVA), outliers were examined per group. Multivariate outliers were defined as having a Cook's distance larger than 1.00 (Tabachnick & Fidell, 2013). Univariate outliers on descriptive variables were not deleted or changed. Univariate or multivariate outliers on variables of interest were individually judged to be one of the following options:

1. If the univariate or multivariate outlier was not part of the sample of interest (e.g., unusual baseline dieting, comments about unique dieting pattern) all responses from the respective participant were deleted from the final sample.
2. If a univariate outlier was part of the sample of interest, the specific value on the offending variable was changed to a raw score that was one unit larger (or smaller) than the next most extreme score in the distribution (Tabachnick & Fidell, 2013). Unless otherwise indicated, the pattern of results did not change because of these transformations.
3. If a multivariate outlier was part of the sample of interest, the analysis was rerun while excluding the outlier. If this changed the pattern of the results, the analysis was reported with the specific notion of deleting the problematic observation. If not, the analysis was reported using the complete sample.

Assumptions

The assumption of normality of univariate distributions was examined by visual inspection of histograms. Assumptions of homogeneity of variance were judged by calculating Hartley's $F_{\max}$, and the assumption was judged to be violated if $F_{\max} > 10$ (Field et al., 2012). Normality and homoscedasticity were assessed by visual examination of diagnostic plots in R and the function for global validation of test assumptions (Kabacoff, 2011). ANOVA was expected to be robust to violations of normality and homogeneity of variance with relatively equal sample sizes (max 1:4), no outliers, and two-tailed tests, and with 20 degrees of freedom for error (Tabachnick & Fidell, 2013). If robustness was not expected, it was assessed whether a transformation of non-normally distributed variables solved the problem of normality. For regression analyses, when assumptions were violated statistical significance was evaluated by calculating bias-corrected bootstrapped 95% confidence intervals.