What Makes for a Good Theory? How to Evaluate a Theory Using the Strength Model of Self-Control as an Example

1

Karolin Gieseler, David D. Loschelder, and Malte Friese

Contents

K. Gieseler · M. Friese (✉)
Saarland University, Saarbrücken, Germany
e-mail: malte.friese@uni-saarland.de

D. D. Loschelder
Leuphana Universität Lüneburg, Lüneburg, Germany

Introduction

The present chapter differs from the others you will read in this book: Chapters in the first section present a specific theory and elaborate on research in applied contexts, in which the respective theory has been used. Chapters in the second section start out with a real-world phenomenon and explain how different psychological theories can help to better understand human behavior or contribute to solving real-world problems. In the present chapter, we take a step back and discuss how the quality of a theory and the quality of its accompanying empirical foundation can be evaluated. In doing so, we distinguish between two different perspectives, the theoretical perspective (Does a theory meet general criteria of a good theory?) and the empirical perspective (How scientifically sound is the research related to a theory?).

In the first part of this chapter, we introduce the Strength Model of Self-Control (Baumeister & Heatherton, 1996; Baumeister & Vohs, 2016). The model will serve as an illustrative reference point throughout the chapter. It is one of the most prominent, researched, and debated theories in social psychology of the last 25 years. Using this concrete example hopefully renders the subsequent discussion of (sometimes) abstract questions and concepts more tangible.

K. Sassenberg, M. L. W. Vliek (eds.), *Social Psychology in Action*,
https://doi.org/10.1007/978-3-030-13788-5_1

The second part of the chapter addresses basic questions relating to scientific theorizing, such as: What is a theory? Why do we need theories? And what makes for a *good* theory? We first discuss criteria for evaluating theories in general before applying them to the Strength Model.

In the third part, we examine criteria to evaluate how theories fare on the *empirical* side. Again, we first discuss criteria for the quality of empirical work in general before applying them to the Strength Model.

The Strength Model of Self-Control

The Strength Model of Self-Control (Baumeister & Heatherton, 1996; Baumeister & Vohs, 2016) originated from observations from everyday life: When reviewing a large and diverse literature, Baumeister and colleagues observed that people who have difficulties following their long-term goals in one life domain often experience similar difficulties in other domains as well. The authors also noticed that self-control failures tend to occur more frequently after long and tiring days, in stressful times, or when demands are unusually high. They concluded that people behave as if **self-control** was a general capacity that is limited and can be depleted.

In their Strength Model, the authors employed the analogy of a (self-control) muscle that becomes tired with use. The model makes two central assumptions: First, self-control draws on a *limited resource;* the exertion of self-control increases the probability of self-control failure in subsequent attempts. Second, self-control is a *domain-general* construct. An exertion of self-control in one domain will increase the likelihood of self-control failure in any other domain that requires self-control. The Strength Model thus assumes a cause-effect relation between the exertion of self-control and the subsequent impairment in self-control performance. Baumeister and colleagues referred to the state of reduced self-control resources as *ego depletion*.

One important implication of the theory's assumptions is that self-control can be improved with practice: If self-control works like a muscle, the repeated exertion of self-control should lead to repeated states of ego depletion, but, in the long run, the muscle should be strengthened, and the overall self-control ability should improve (e.g., Job, Friese, & Bernecker, 2015; Muraven, 2010; for meta-analyses, see Beames, Schofield, & Denson, 2017; Friese, Frankenbach, Job, & Loschelder, 2017). In the present chapter, we focus on those aspects of the Strength Model that are concerned with the **ego depletion effect** rather than with the trainability hypothesis.

Definition Box

Self-Control: "Ability to override or change one's inner responses, as well as to interrupt undesired behavioral tendencies (such as impulses) and refrain from acting on them" (Tangney, Baumeister, & Boone, 2004, p. 274).

Ego Depletion Effect: A person shows impaired performance in self-control demanding tasks after she has previously exerted self-control (compared to a control group that has not exerted self-control in task 1).

To test the model's core assumptions in the psychological laboratory, Baumeister and colleagues developed the *sequential task paradigm*: They had participants work on two sequential tasks demanding self-control and measured their performance in the second task as a function of whether the first task was high or low in self-control demands (e.g., Baumeister, Bratlavsky, Muraven, & Tice, 1998). Hundreds of studies following this paradigm have provided evidence for this ego depletion effect (for reviews and meta-analyses, see Baumeister & Vohs, 2016; Carter, Kofler, Forster, & McCullough, 2015; Hagger, Wood, Stiff, & Chatzisarantis, 2010; Hirt, Clarkson, Egan, & Eyink, 2016). For example, Muraven, Collins, and Neinhaus (2002) had participants either suppress their thoughts in a first task (which requires self-control) or not. Those who exerted self-control in the first task consumed

more beer during a subsequent taste test, even though participants knew they were about to do a driving test afterward. In another exemplary study, chronic dieters who suppressed their emotional reactions to a sad video (a self-control demanding task) later consumed more ice cream during a product test than those who were instructed to react naturally while watching the video (Vohs & Heatherton, 2000).

Box 1.1 Questions for Elaboration

1. Assume the Strength Model of Self-Control is accurate. Think about activities in your daily life that should lead to depletion-like effects.
2. Think about possible strategies that you could use to counteract ego depletion effects in your daily life. These could be strategies that you have used yourself or that you may know from other chapters in this book (e.g., Rubicon model and implementation intentions, Keller et al., Chap. 2; Mindset theory, Bernecker & Job, Chap. 12).
3. Can you think of situations when your self-control felt unlimited? How did these situations differ from when you felt depleted, lacking ability to self-control?
4. Assume the Strength Model of Self-Control was *in*accurate: The ability to exert self-control was not limited. What would your daily life and the world more generally look like? Any different from the present reality?

Despite the seemingly abundant evidence in favor of the Strength Model, the model and its accompanying empirical work have been heavily criticized in recent years. These criticisms go so far that many researchers doubt that the ego depletion effect is a real phenomenon after all. We will elaborate on some of these issues that have been criticized in later parts of this chapter (for an overview of the debate, see Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2019).

Behavioral Versus Process Level of Psychological Phenomena

Before we turn to the discussion of criteria to evaluate a theory, we need to introduce an important distinction between two different levels of analysis that will guide our further thinking: the distinction between the *behavioral* and the *process level* (also referred to as "functional" and "cognitive" level of analysis, see De Houwer, 2011; Fig. 1.1). The **behavioral level of analysis** defines behavioral effects exclusively in terms of changes in elements of the environment that cause behavioral changes on a dependent variable. For instance, insulting someone increases aggressive behavior in the insulted person. By contrast, the **process level of analysis** refers to the underlying mental processes that are triggered by elements in the environment and are responsible for subsequent changes on a dependent variable. For example, an insult may trigger anger that then translates into aggressive behavior. These two levels of analysis must not be conflated. In the words of De Houwer (2011, p. 201):

> "… using behavioral effects as a proxy for mental constructs violates the general scientific principle that the *explanandum* (that which needs to be explained; in this case, behavioral effects) needs to be kept separate from the *explanans* (that which is used to explain; in this case, mental constructs; Hempel, 1970)."

Definition Box

Behavioral level of analysis: Defining behavioral effects exclusively in terms of elements in the environment.

→ Which elements in the environment lead to a certain behavior? (De Houwer, 2011)

Process level of analysis: Examining the nature of underlying mental processes that are assumed to guide behavior/behavioral effects.

→ Via which underlying mental process(es) do certain elements in the environment lead to a certain behavior? (De Houwer, 2011)

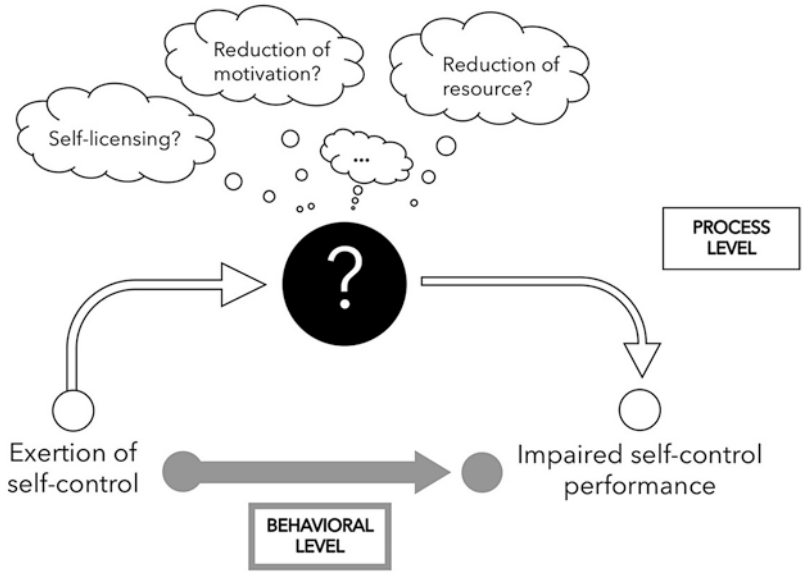

Fig. 1.1 Levels of scientific analysis for the example of the ego depletion effect. The behavioral level denotes the cause-effect relation: The exertion of self-control leads to subsequent impairments in self-control performance (gray arrow). The process level refers to the underlying psychological mechanisms that are triggered by the exertion of self-control—that is, the mechanisms on the process level are causally responsible for the behavioral effect. (White arrows via question mark; adapted from De Houwer, 2011)

The Strength Model makes assumptions on both levels of analysis. On the behavioral level, it states that the initial exertion of self-control causes a subsequent impairment in self-control performance, linking two elements in the environment: self-control exertion at time one and impaired self-control performance at time two (relative to a control group). On the process level, the model assumes that the behavioral effect is mediated by the depletion of an internal, limited resource (the nature of this resource is not further defined). For present purposes, the distinction is important as there are several other theoretical models beyond the Strength Model seeking to explain the same behavioral phenomenon (behavioral level) with fundamentally different assumptions concerning the underlying process (process level; for further reading see, e.g., De Witt Huberts, Evers, & De Ridder, 2014; Evans, Boggero, & Segerstrom, 2015; Inzlicht, Schmeichel, & Macrae, 2014).

Theoretical Perspective: Criteria to Evaluate the Quality of Theories

Having briefly described the Strength Model's basic assumptions and some accompanying empirical evidence, let us take a step back: the empirical foundation aside, is the Strength Model a "good" theory to start with? And more generally, what makes for a good theory? In science, a theory constitutes one or several joined-up principles that are meant to describe, explain, and predict a phenomenon or several related phenomena (Estrada & Schultz, 2017). A theory is not necessarily true. It may be (partly) true, but it may also be (partly) false. A theory is a set of ideas meant to explain observable events. Appropriate scientific methods are needed to test whether or not a theory achieves this aim. Theories thus are the basis to expand our understanding of the world. For social psychologists,

they are the starting point for interventions to address individual and social problems and to change problematic behavior (see second section of this book: Combining theoretical insights: Addressing complex human behavior).

Taking the Strength Model as an example, one could say that the model's two main assumptions—*limited resource* and *domain-general construct*—are meant to predict and explain the phenomenon of impaired self-control performance after the exertion of self-control. This pertains to the psychological laboratory *and* to people's everyday life. In the long run, if the model stands the test of time, interventions based on the Strength Model may thus address self-control failures across various domains often challenging self-control such as eating, drinking, exercising, social interactions, and procrastination, among others.

Not all theories are *good* theories, however. And to distinguish the good from the not so good, there are several criteria to consider. Here, we focus on six criteria, namely, consistency, precision, parsimony, generality, falsifiability, and progress, while omitting (partly overlapping) criteria such as refutability or truth (see Gawronski & Bodenhausen, 2015; Van Lange, 2013; see Table 1.1). Due to space restrictions, our list and discussion are necessarily incomplete. A more in-depth treatment can be found in Gawronski and Bodenhausen (2015), our primary source for this part of the chapter. We start each of the following sections by first defining the respective criterion and then subsequently applying the criterion to the Strength Model.

Consistency

One obvious characteristic of a good theory is consistency with empirical observations. If a theory does not correspond to empirical observations in the laboratory and/or the real world, it is necessary to adjust the theory (or to refute it). Sometimes a theory turns out to have merit only after some conceptual adjustment. For example, research may identify boundary conditions that specify when predictions derived from the theory do or do not apply. If after adjustments a theory still is not consistent with empirical observations to a satisfactory extent, it may be necessary to abandon the theory.

The Strength Model originates from the observation in the psychological literature that in everyday life people seem more likely to fail at controlling themselves after previously exerting self-control. An *in*consistent observation is that there seem to be other situations in everyday life in which people appear to have no difficulty to

Table 1.1 Selection of quality criteria that make for a good theory and their application to the Strength Model

Criterion	Definition	Application to the Strength Model
Consistency	Correspondence to empirical observations in the laboratory and/or the real world	Hundreds of lab studies and real-world observations consistent with phenomenon. Inconsistent recent (large-scale) replications and preregistered studies
Precision	Clearly defined concepts and operationalizations that allow for little stretching	Imprecise definition of self-control and the limited self-control resource
Parsimony	Explain more with less: Use as few assumptions as possible to explain a given phenomenon	Two core assumptions—limited resource that is domain independent—account for a far-reaching phenomenon
Generality	Favor higher explanatory breadth	Model's assumptions apply to and are observable in a large array of situations, contexts, and behaviors
Falsifiability	Formulate assumptions so that it is possible to make observations prohibited by a theory	Imprecise formulation of self-control and underlying resource make it difficult to falsify some of the theory's predictions
Progress	Inspire new research and discoveries and promote theoretical progress	Theory has spurred hundreds of studies, novel theorizing, and methodological, scientific debates

Inspired by Gawronski and Bodenhausen (2015)

resist their temptations and to successfully control their impulses—even after a demanding previous task. These observations alone do not mean that the Strength Model has to be abandoned. Indeed, several situational and dispositional moderators have been identified that presumably prevent or counteract ego depletion effects (for a review, see Loschelder & Friese, 2016). For example, affirming core personal values (Schmeichel & Vohs, 2009) or being incentivized to perform well have been shown to counteract ego depletion effects (Luethi et al., 2016; Muraven & Slessareva, 2003). In a similar vein, holding a subjective theory that self-control is non-limited (Job, Dweck, & Walton, 2010; Bernecker & Job, Chap. 12) or having a high disposition for action orientation (Gröpel, Baumeister, & Beckmann, 2014) has been found to prevent the occurrence of ego depletion. In all, boundary conditions are crucial and need to be defined well in order to account for theory-consistent and inconsistent findings.

Another (potentially greater) problem for the Strength Model stems from the increasing number of studies that fail to find ego depletion effects without moderating variables being able to explain these inconsistent data (e.g., Etherton et al., 2018; Lurquin et al., 2016; Osgood, 2017; Singh & Göritz, 2018; Vadillo, Gold, & Osman, 2018). The empirical evidence does not seem to as consistently support the theoretical assumptions as was believed for many years. From our perspective, conceptual and empirical work is necessary to address this lack of consistency (especially in light of doubts about the ability of earlier empirical work on the model to lead to firm conclusions, see below and Friese et al., in press). Otherwise, a lack of consistency will seriously threaten the state of the Strength Model as a respected theory.

Precision

A good theory is *precise*, with clearly defined concepts and operationalizations that allow for little stretching or subjective interpretation. The more precise the formulation of a theory and its background assumptions, the less ambiguous it is for researchers to decide which empirical observations are consistent versus inconsistent with the theory (Gawronski & Bodenhausen, 2015). Thus, precision increases the chances of collecting both supporting and refuting empirical evidence for a theory. Imprecise theories leave room for subjective interpretation of empirical findings.

Precision is not a strength of the Strength Model. One problem of the theory is shared with the field of self-control research in general: a precise (and widely accepted) definition of self-control is lacking. Baumeister and Vohs (2016, p. 70) define self-control "as processes by which the self intentionally alters its own responses, including thoughts, emotions, impulses, performance, and behaviors, based on standards." This definition encompasses a great part of what people intentionally do when awake. But not every time a person intentionally alters her own responses to be in line with her standards, she will exert effortful self-control leading to depletion effects. For example, picture a person writing an official Email in line with orthography and grammar. Writing requires altering one's responses based on standards (the norms of orthography and grammar). But for someone educated enough to write decently, this is so low-minded that it seems implausible to assume that such a task will easily lead to discernable depletion-type effects. Thus, this definition of self-control may be too unspecific and likely too broad. Similar points can be made with respect to the alternative definition of self-control provided in the Definition box earlier in this chapter. In all fairness, somewhat vague and imprecise conceptual definitions are nothing unique to the Strength Model, but a feature shared with many other theories and research fields in (social) psychology.

The *im*precise definition of self-control is relevant for the antecedent of impaired self-control (i.e., the independent variable in experimental studies). Here, it pertains to the behavioral level of analysis: What exactly are properties of a valid manipulation that allow for a stringent test of the model, independent of the behavioral effect it may or may not evoke?

The Strength Model's lack of precision becomes especially apparent on the process level: The model postulates the reduction of a "limited resource" as the crucial underlying mechanism that is causally responsible for impaired self-control performance. However, the model does not specify the nature of this resource in any way. In consequence, the model remains too imprecise to test one of its core assumptions as it is impossible to measure an unknown resource.[1]

The lack of precision on the behavioral and process level may contribute to the problems and reproaches that ego depletion research is facing: If researchers do not precisely know how to manipulate the exertion of self-control, it is difficult to distinguish a nonsignificant result caused by a wrong theory from a nonsignificant result caused by an unapt experimental manipulation. Similarly, it is impossible to provide conclusive evidence supporting the underlying process assumptions, if the model does not provide clear guidance about the properties of this very resource (or how to measure it).

Parsimony

A good theory explains more with less. When developing a theory, one possibility is to start with observations, to abstract and generalize them, and to create joined-up principles (see definition above). In a next step, the theorist cuts out all superfluous elements. A straightforward formula remains that predicts a multitude of events with as few assumptions as possible. Hence, when comparing two theories that both explain the same set of empirical observations, the theory with fewer assumptions is superior in terms of parsimony.

The Strength Model is very parsimonious as it originally made only two central assumptions (domain generality and limited resource). Many (social) psychological theories make many more than just two central assumptions (e.g., Social Identity Theory, Scheepers & Ellemers, Chap. 9; Social Learning Theory, Bandura, 1977; General Aggression Model, Bushman & Anderson, 2002). Excessive parsimony comes with the downside that it may impair consistency—thus eventually requiring conceptual additions to a theory. Future research will reveal if failures to detect the ego depletion effect can be accounted for by such boundary conditions (moderators) and auxiliary assumptions—at the cost of parsimony. (Alternatively, failures to find ego depletion effects may also be due to problems discussed later in the section "Empirical perspective").

Generality

Generality refers to a theory's quality to apply to various fields, situations, or domains of behavior. Although theoretical parsimony is an asset of a theory, explanatory breadth is as well.

The Strength Model fares very well concerning this criterion: The proposed explanatory breadth of the model is unusually large. A wide array of behaviors in almost all spheres of life require self-control and may—if the theory is right—evoke ego depletion effects. Likewise, according to the Strength Model, exerting self-control may have an impact on a multitude of behaviors in a variety of areas, all of which therefore should be susceptible to ego depletion effects.

Falsifiability

No matter how much empirical evidence has been gathered that is consistent with a theory, the theory can never be inductively "proven" to be true—it is always possible that one day an observation inconsistent with the theory emerges (Popper, 1959). All the more important, a theory has to be falsifiable. If a theory is formulated in a way that makes it impossible to observe something that is prohibited by the theory's assumptions, it is unfalsifiable and therefore not a good

[1] Some earlier studies had suggested that blood glucose levels may be this underlying limited resource, but this idea has been dismissed on both logical, physiological, and empirical grounds (Dang, 2016; Kurzban, 2010; Vadillo, Gold, & Osman, 2016; but see Ampel, Muraven, & McNay, 2018, for a different perspective).

theory (e.g., the claim "the exertion of self-control subsequently leads to better, poorer or unchanged self-control performance" would make the theory unfalsifiable). It is, of course, possible to define already mentioned auxiliary assumptions or boundary conditions to explain observations that initially appear inconsistent with the theory. However, every inconsistent observation must not lead to the development of a new auxiliary assumption specifying a new exception. Especially if there are more exceptions to the theory than standard cases, a theory becomes unfalsifiable (see also criterion of Parsimony).

The Strength Model states that the exertion of self-control impairs self-control performance. This claim would be falsified, if the exertion of self-control was consistently found to *boost* rather than to impair performance. Savani and Job (2017) found such a reverse ego depletion effect in several studies. Importantly, however, participants in these samples came from India growing up with the cultural belief that exerting self-control is *beneficial* for future self-control exertion. Adding "cultural belief" (or beliefs about the [non-]limitedness of self-control more generally) as a boundary condition incorporates these results into the larger theory.

Some other assumptions of the Strength Model, however, seem difficult to falsify. On the behavioral level, the prediction that exerting self-control impairs subsequent self-control performance is difficult to falsify because due to the imprecision of the self-control definition, it is unclear what constitutes a valid self-control manipulation and a valid dependent variable. Failures to replicate an effect can easily be dismissed by doubting the validity or the strength of the independent or dependent variables (see section on Operationalization and Manipulation Checks below). On the process level, the nature of the purported resource is unspecified, as mentioned earlier. Its existence can therefore not be falsified.

Progress

Good theories inspire new research, lead to discoveries that make contributions beyond the previously known, and promote theoretical progress through refinements, sharpening, and the inspiration and development of (new) theories.

The Strength Model has successfully spurred progress. Hundreds of studies have been conducted and endorsed the existence of the ego depletion effect (Hirt, Clarkson, & Jia, 2016). The model has been applied to many different spheres of psychology including consumer behavior (Friese, Hofmann, & Wänke, 2008; Vohs & Faber, 2007), neuroscience (Heatherton & Wagner, 2011; Luethi et al., 2016), decision-making (Pocheptsova, Amir, Dhar, & Baumeister, 2009), and work and organizations (Christian & Ellis, 2011), to name a few. The model also inspired the development of new theories explaining the ego depletion effect differently (e.g., Central Governor Model: Evans et al., 2015; Process Model: Inzlicht et al., 2014) and more general theories of self-control integrating ego depletion as a central component (e.g., De Witt Huberts et al., 2014; Kotabe & Hofmann, 2015). Furthermore, methodological discussions triggered by doubts about the robustness of the ego depletion effect led to advances in domains such as research on publication bias (Inzlicht, Gervais, & Berkman, 2015). In sum, the Strength Model was extraordinarily successful in stimulating both empirical work and theory across different fields in psychology and beyond. (Obviously, in and of itself this does not make a theory any more or less true.)

Interim Summary

In the first part of this chapter, we introduced the Strength Model of Self-Control with its two main assumptions: domain generality and limited resource. We discussed the sequential task paradigm and introduced two levels of analysis: the behavioral level and the process level. We proceeded with a selection of criteria to evaluate the quality of theories and applied these to the Strength Model. In the following section of the chapter, we turn to the examination of the quality of empirical research that has been conducted to test a theory, again using the Strength Model as an illustrative example.

Box 1.2 Questions for Elaboration

1. Think of a social psychological theory you know, and try to evaluate it along the criteria presented in this part of the chapter. In what respects is the theory you chose a good theory? Where is room for improvement? Discuss.
2. If you were to rank the discussed criteria, which of these would you see as the more relevant indicators for a high-quality theory? Are some of these criteria mutually exclusive?

Empirical Perspective: Criteria to Evaluate Research on a Theory

We now take a look at a selection of criteria that help to judge the quality of empirical research that has been conducted on a theory. Admittedly, these criteria are inspired specifically by discussions about the Strength Model. They are therefore neither exhaustive nor representative for evaluating empirical work in general. Nevertheless, several of these criteria can readily be applied to evaluating research on other theories as well. For a more in-depth discussion of empirical work on the ego depletion effect, see Friese et al. (in press).

Statistical Power

We start with the famous Jacob Cohen (1988, p. 1): "The power of a statistical test is the probability that it will yield statistically significant results [for an effect that truly exists]." To reliably detect an effect that truly exists, high statistical power is vital. Importantly, statistical power increases with larger effect sizes and with larger sample sizes (given the significance level for a type-I error is held constant). Thus, if statistical power is low, a study is less likely to detect a true effect. But—and maybe less intuitively—low statistical power also decreases the likelihood that a statistically significant finding reflects a true effect (Button et al., 2013; Christley, 2010). This so-called positive predictive value is lower for smaller effect sizes and for smaller sample sizes.

In the ego depletion literature, many studies have small sample sizes (Carter et al., 2015). Thus, in combination with a true effect which—by now—is assumed to be small in size, statistical power across the ego depletion literature is assumed to be worryingly low. When the average power in a literature is low, many studies are likely to produce nonsignificant findings even in the presence of a true effect. However, the vast majority of published ego depletion studies reveal significant effects (Carter et al., 2015). Together, these observations limit the possibility to draw firm conclusions concerning the Strength Model based on the currently available evidence (see also later section on *P*-Hacking and Publication Bias).

Operationalization and Manipulation Check

The process of defining an instrument to measure a phenomenon that is not directly observable is called operationalization—the resulting representations of the phenomenon are *operational definitions*. For instance, to measure the abstract concept of a person's intelligence (which is not directly observable), psychologists have developed many different intelligence tests that seek to measure and quantify the underlying construct. In principle, a phenomenon can have an unlimited number of operational definitions (Whitley, 2002). It is important to thoroughly develop these operational definitions because researchers draw inferences concerning the phenomena (i.e., hypothetical, latent constructs; e.g., intelligence) based on the measurement of observable, manifest variables (e.g., scores in an intelligence test). Only if this relationship between manifest variables and the phenomenon is trusted—if the operational definition fits the theory—can we draw conclusions concerning the hypothetical construct. Otherwise, if central constructs are not well operationalized, it is difficult to estimate the robustness of an effect.

Research on the Strength Model used a large variety of operational definitions concerning the measurement and the manipulation of self-control (see section on Generality above). The Strength Model itself does not explicitly suggest a certain set of operational definitions, partly due to its generality (see Theoretical Perspective). Hence, there is a large variety of ego depletion manipulations in terms of time and content: from very brief manipulations such as 20 incongruent (depletion condition) versus 20 congruent (control condition) Stroop trials (Yam, Chen, & Reynolds, 2014) to completing several demanding tasks in a row in the depletion condition, each lasting several minutes (Sjåstad & Baumeister, 2018; Vohs, Baumeister, & Schmeichel, 2012), and from resisting cookies versus radishes to crossing out certain letters in a text (Baumeister et al., 1998). The same is true for the dependent variables: A large variety of different tasks have been used as dependent variable, ranging from executive functioning tasks like the Stroop task (Inzlicht & Gutsell, 2007) to arithmetic calculations (Vohs et al., 2008) via tasks involving resisting temptations (Christiansen, Cole, & Field, 2012; Friese et al., 2008) to risk taking (Freeman & Muraven, 2010) and aggressive behavior (DeWall, Baumeister, Stillman, & Gailliot, 2007). Meta-analytic evidence suggests that these tasks share a common core, albeit a rather small one (Duckworth & Kern, 2011). Thus, the question is, if and to what extent these operational definitions can reliably and validly manipulate and measure the underlying construct that the Strength Model posits: self-control.

A related point concerns the use of manipulation checks. Manipulation checks are typically used to test whether an independent variable successfully manipulated the construct of interest. The majority of ego depletion studies did not make use of manipulation checks. In a first meta-analysis of 198 studies (Hagger et al., 2010), only 30% of those studies included a manipulation check asking for perceived difficulty of the first task. Out of these 30%, only about half (15.7% in total) asked for subjective effort, and only a little more than one in ten (12.6%) assessed fatigue after the first task. This is important to keep in mind: When researchers have little evidence whether and to what extent participants exerted self-control, the success of the manipulation may be in doubt. This, in turn, has implications for the falsifiability of the model: A failure to find an effect could be due to problems from the theoretical perspective (i.e., the model's assumptions may not be correct) or from the empirical perspective (i.e., the model is correct, but the unsuccessful manipulation of the relevant constructs did not allow for a proper test). In the latter case, it may be premature to dismiss a theoretical model, even in light of nonsignificant findings.

p-Hacking and Publication Bias

In recent years, several issues have been debated that may contribute to less-than-desirable replicability and robustness of psychological science (e.g., Munafò et al., 2017). Two sources of bias in particular have received widespread attention: *p*-hacking and publication bias.

p-hacking refers to researchers engaging in questionable research practices to make originally nonsignificant analyses statistically significant. A nonsignificant *p*-value is "hacked" to significance (see Fig. 1.2). Consequently, findings appear more robust than they actually are (Simmons, Nelson, & Simonsohn, 2011, 2018). Common *p*-hacks include reporting only dependent variables that "worked" while omitting others, including or excluding outliers depending on which analyses reveal the more desirable outcome, peeking at the data during data collection and stopping when the desired pattern of results emerges without controlling for the increased Type-I error rate (α error, probability of rejecting the null hypothesis when it is in fact true), or including covariates without a predefined theoretical rationale. Some of these *p*-hacks are especially "efficient" in changing results in small samples.[2]

The most tangible consequence of *p*-hacking is an increase of significant findings that would not

[2]See http://shinyapps.org/apps/p-hacker/ for a vivid demonstration (Schönbrodt, 2015).

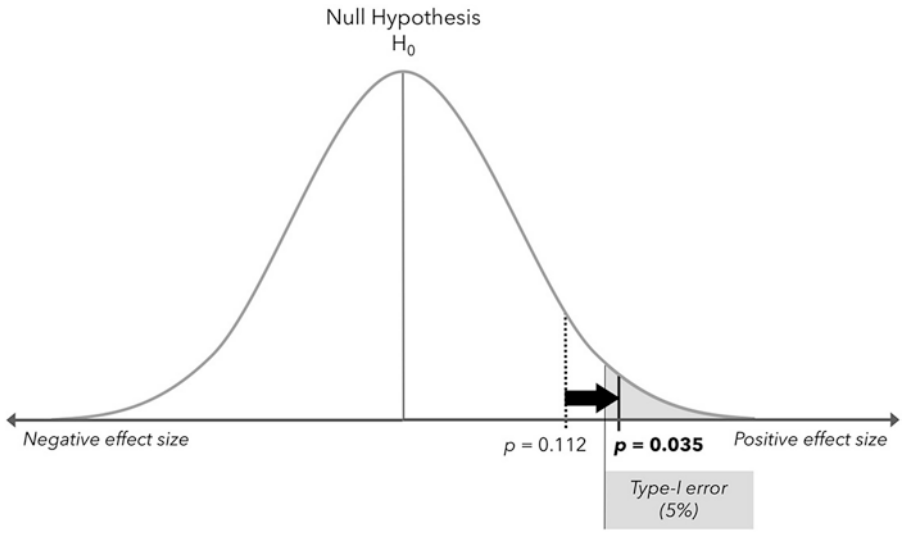

Fig. 1.2 An illustrative example of p-hacking and its consequences. A nonsignificant effect with $p = 0.112$ (dotted line) is *p*-hacked (black arrow) below the common 0.05 significance threshold to $p = 0.035$ (continuous line). As a consequence, the size of the effect increases. However, the main objective in *p*-hacking is to obtain a significant result, not to increase the effect size

have been significant without *p*-hacking. An original *p*-value may be nonsignificant while showing a (nonsignificant) tendency in the expected direction. After *p*-hacking, the effect size is artificially inflated, and the finding is significant. These false-positive findings due to *p*-hacking suggest the presence of a true effect that in fact may not exist or may be smaller than suggested. One important consequence of *p*-hacking is that it may lead to an inflation of significant findings in a given literature that may contribute to convictions about the existence of a phenomenon that in fact may be much less reliable than it appears. Another important consequence of *p*-hacking is that it may lead to an overestimation of the effect size for a given phenomenon that may in fact be much smaller than it appears.

The pervasiveness and severity of *p*-hacking is unknown and estimates vary widely (Fiedler & Schwarz, 2016; Hartgerink, 2017; Head, Holman, Lanfear, Kahn, & Jennions, 2015; John, Loewenstein, & Prelec, 2012). It seems safe to say that *p*-hacking has played a role in the ego depletion literature (Wolff, Baumann, & Englert, 2018), as it has in many other psychological literatures. *P*-hacking will have contributed to the overestimation of meta-analytic effect size estimates in ego depletion research, although it is unclear to what extent this is the case (Friese et al., in press). In addition, *p*-hacking likely has produced a number of published findings that would not have been significant without *p*-hacking (and therefore less likely to be published).

Publication Bias Publication bias refers to the observation that studies with statistically significant results are more likely to be published than nonsignificant studies (Bakker, van Dijk, & Wicherts, 2012). The most tangible consequence of publication bias is the overestimation of meta-analytic effect size estimates, because the number of published studies (with often larger effect sizes; see above) is not adequately corrected downward by existing, but unavailable nonsignificant studies with *smaller* effect sizes. Thus, publication bias can lead to a distorted perception of the magnitude and robustness of research literatures.

Estimating the severity of publication bias is difficult, but some analysts suggest that it is generally high in the social sciences including psychology (Fanelli, 2010; Fanelli, Costas, & Ioannidis, 2017) and also for ego depletion research in particular (Carter et al., 2015; Carter & McCullough, 2014). This is troublesome,

because common techniques to correct for the influence of publication bias work poorly under conditions typical for psychological science (e.g., heterogeneity in effect sizes; Carter, Schönbrodt, Gervais, & Hilgard, 2017).

Taken together, *p*-hacking (particularly) increases the rates of false positives, while publication bias (particularly) increases meta-analytic effect size estimates. Based on the preponderance of statistically significant findings despite low power, it is plausible to assume that both *p*-hacking and publication bias have contributed markedly to ego depletion research. Together, they convey the impression of a more robust and replicable literature with larger effect sizes than is warranted. How severe exactly their influence is, is unfortunately impossible to determine. Studies using open science practices such as preregistration (Nosek, Ebersole, DeHaven, & Mellor, 2018; van 't Veer & Giner-Sorolla, 2016), open materials, and open data are less prone to the deleterious effects of *p*-hacking and publication bias (Munafò et al., 2017). It should therefore become a habit for researchers to preregister their predictions on an openly accessible online forum, where they can also share their experimental materials, original data, and analysis scripts. Future work (not only on ego depletion) should embrace open science practices.

Moderation and Mediation

Moderators can reveal important boundary conditions of effects proposed by a theory (see first part of this chapter). They can elucidate the breadth of a phenomenon, reveal new differentiations, and explore limits of a theory.

More than 100 studies have investigated moderators of the ego depletion effect (for an overview, see Loschelder & Friese, 2016). For example, incentives to perform well in a second self-control task can counteract ego depletion effects (Luethi et al., 2016, see Fig. 1.3a; Muraven & Slessareva, 2003), and people who believe that their willpower is non-limited may

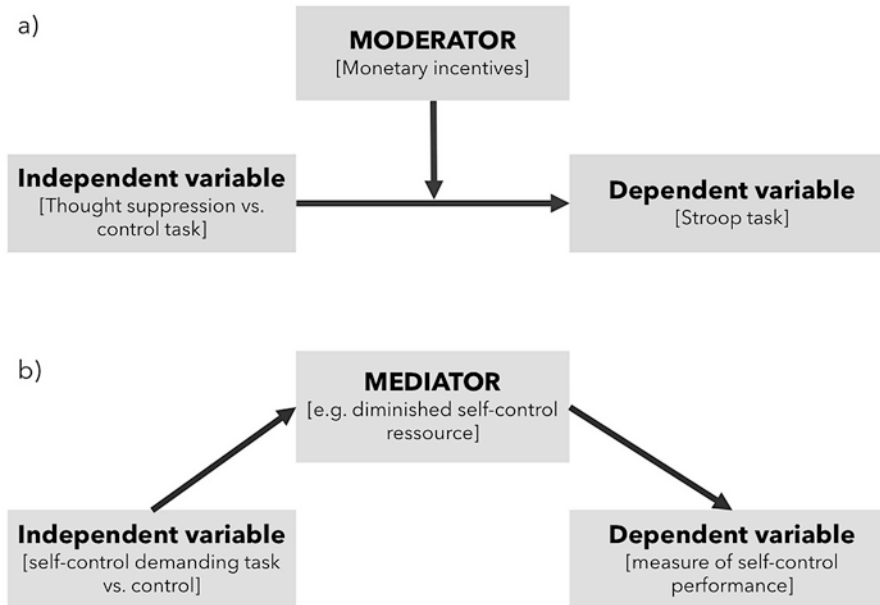

Fig. 1.3 (**a**) Moderation: The relationship between independent and dependent variable is influenced by a third variable, the moderator. Luethi et al. (2016), for example, found that the effect of the self-control manipulation on the performance in a Stroop task depended on whether participants could earn additional money depending on their performance in the Stroop task or not. (**b**) Mediation: The relationship between independent and dependent variable—the underlying process—can (partly) be explained by a third variable, the mediator. The Strength Model, for instance, assumes that the diminution of the self-control resource explains why after exerting self-control in a first task, people's self-control performance is impaired in a second self-control demanding task

be less likely to show ego depletion effects (Job et al., 2010; Bernecker & Job, Chap. 12). One may suspect that moderator studies (that are often investigated in 2 × 2 experimental designs) are more difficult to *p*-hack than two-condition studies and may therefore provide more robust results. However, this assumption is unlikely given that only one experimental condition would need to be *p*-hacked for a moderation pattern (Friese et al., in press). In addition, moderator studies in the ego depletion literature predominantly report significant findings, but had low statistical power, suggesting the presence of *p*-hacking and/or publication bias.

Mediation studies are used to test assumptions about the underlying process of a phenomenon (see process level of analysis, De Houwer, 2011). Mediators can therefore—in principle—distinguish between different theoretical explanations of the same observations. Mediators can be measured or manipulated (e.g., Hayes, 2013; Spencer, Zanna, & Fong, 2005). When a proposed mediator is measured, the data pattern necessary to obtain significant mediation is more complex than a mean difference between two conditions. Mediator studies may thus—again, in principle—be less likely to produce false positives and instead provide more robust evidence for a phenomenon.

While providing process evidence for the Strength Model (i.e., diminished self-control resource; Fig. 1.3b) is impossible because the assumed resource is unspecified and thus nonassessable, some researchers examined other mediators not specified by the Strength Model (e.g., Chow, Hui, & Lau, 2015; Graham, Martin Ginis, & Bray, 2017; Inzlicht & Gutsell, 2007). Considering the size of the ego depletion literature, mediator studies are rare. They commonly have low statistical power, too. Importantly, since hardly any studies on ego depletion were preregistered, it is impossible to know how many mediator studies were conducted but not reported at all or in a different manner because the mediation results did not turn out as expected. The existing evidence in the ego depletion literature is therefore limited. More generally, we believe that preregistered and theoretically grounded mediator studies are capable of providing stronger evidence for a theory as they examine an assumed mechanism on the process level in addition to the behavioral effect.

Meta-Analyses

Meta-analyses are a powerful tool to combine the results of multiple relevant studies in a research field (Borenstein, Hedges, Higgins, & Rothstein, 2011). They shift the focus from individual studies to the broader picture. Some strengths of meta-analyses are the higher statistical power to show even small effect sizes and the ability to examine moderators across studies that are unfeasible to investigate in individual studies. Despite their many benefits, meta-analyses have drawbacks. An important one is that the quality of a meta-analysis crucially depends on the quality of the primary studies (Egger, Smith, & Sterne, 2001; Ioannidis & Lau, 1998). If a field features many poorly conducted studies, a meta-analysis will unlikely level out the biases of primary studies (e.g., *p*-hacking; see also Munafò et al., 2017)—particularly, if these biases are systematic rather than unsystematic (Borenstein et al., 2011). As mentioned earlier, publication bias can also distort meta-analytic effect size estimates. The possibilities to reliably correct for publication bias are limited (Carter et al., 2017). Nevertheless, meta-analyses currently are the most potent way to quantitatively summarize large sets of studies (Gurevitch, Koricheva, Nakagawa, & Stewart, 2018).

For the ego depletion literature, a first meta-analysis of published studies indicated a healthy mean effect size of $d = 0.62$ (Hagger et al., 2010). However, a reanalysis of the same dataset found evidence for publication bias (Carter & McCullough, 2014). A second meta-analysis found an uncorrected smaller, but still substantial effect of $g = 0.43$ (Carter et al., 2015). After applying different techniques to correct for the influence of publication bias (that all have some shortcomings), Carter and colleagues concluded that there is "very little evidence that the ego depletion effect is a real phenomenon" (Carter et al., 2015, p. 796).

Some researchers saw this meta-analysis by Carter et al. (2015) as the first nail in the coffin of ego depletion research. Others pointed out various problems of the meta-analysis, the bias correction methods, and questioned its conclusions (Cunningham & Baumeister, 2016; Inzlicht et al., 2015). In the meantime, further meta-analyses appeared with varying results that are difficult to interpret due to methodological issues (Blázquez, Botella, & Suero, 2017; Dang, Liu, Liu, & Mao, 2017; see Friese et al., in press). Overall, the inconvenient truth is that ego depletion meta-analyses served as a great tool to promote discussion and progress, but they did not provide unequivocal evidence for (or against) the ego depletion effect.

Registered Replication Reports

Registered Replication Reports (RRR) are "multi-lab, high-quality replications of important experiments in psychological science along with comments by the authors of the original studies" (Association for Psychological Science, 2018). A detailed description of the study, the hypotheses, and the analysis plan is implemented by several laboratories, and the results are published independent of the results. An RRR has thus great statistical power to test a central prediction of a theory. A limitation of RRRs is that they are (usually) restricted to replicating one specific, often prototypical operationalization in a research field (a selected "landmark study"). The ability of RRRs to speak to the validity of whole theories is therefore necessarily limited.

An ego depletion RRR (Hagger et al., 2016) sought to replicate one selected combination of manipulation and dependent variable (Sripada, Kessler, & Jonides, 2014). As we have seen in the first part of the chapter, a salient characteristic of the Strength Model is its domain generality assumption. Thus, the ability of this (or any other) specific IV-DV combination alone to disprove the general ego depletion idea is limited. That being said, the RRR delivered a null effect on average. Although there was discussion about some methodological issues of the RRR (Arber et al., 2017; Baumeister & Vohs, 2016), this finding posed another serious threat to ego depletion research.

Discussion

In this chapter, we discussed criteria to evaluate the quality of theories and the empirical research examining these theories. Instead of reiterating our conclusions, we wish to briefly reflect about (1) the choice to abandon theories, (2) the two levels of analysis in general, and (3) the Strength Model in particular.

A psychological theory is of inferential nature—it makes probabilistic, imperfect predictions of the future—and, as such, can never be *proven* to be true. The process of developing a psychological theory is hence never completed. It is an ongoing journey of testing, refinement, development, falsification, and sometimes rejection. As we tried to show in this chapter, it is important to distinguish between the theoretical perspective and the empirical perspective when working with theories. Failures to replicate an effect (or find it in the first place) can be due to different causes: the theory could be wrong and the original result was a false positive. Similarly, however, the theory could have merit, but the empirical research was not good enough to adequately test it. In this case, one should take a close look at the complete process of putting the theoretical prediction to a practical test. Did the theory make plausible assumptions? Were operationalizations adequate? Such discussion may lead to important insights and sensible further steps for a research field. It may be more promising than a (possibly) premature decision to abandon a theory altogether. From the theoretical perspective, theories may be true after all, even if some empirical tests did not deliver the expected results.

When discussing both the criteria to evaluate the quality of theories and its accompanying empirical research, the Strength Model fared better on some criteria than on others. Here, we wish

to stress that we chose the Strength Model as an illustrative example—not because we sought to promote or undermine this model—but because it has received considerable attention in the last decades (and we happen to have gained some knowledge on it). In addition, many of the doubts and criticisms that we discussed with respect to the Strength Model are not unique to this model; they generalize to several other literatures in social psychology and beyond. The discussion about the scientific implications of these observations for (social) psychology more generally goes past the scope of this chapter.

Summary

- Theories can be discussed from different perspectives: the theoretical perspective (Does a theory meet general requirements for a good theory?) and the empirical perspective (How empirically proven is a theory?).
- A good theory in the theoretical sense is (1) consistent with empirical observations; is (2) precise, (3) parsimonious, (4) explanatorily broad, and (5) falsifiable; and (6) promotes scientific progress (among others; Table 1.1).
- To convincingly support a theory's assumptions, empirical research has to be high in statistical power, well operationalized, and (largely) free of *p*-hacking and publication bias. Meta-analyses and Registered Replication Reports are useful tools to estimate effect sizes, examine moderators, and test central predictions with high statistical power.
- The Strength Model of Self-Control explains self-control failure after the initial exertion of self-control with a domain-general, limited resource. Concerning the criteria for a good theory and well-conducted empirical research, the Strength Model has both favorable and less favorable properties.

Recommended Reading

Baumeister, R. F., & Vohs, K. D. (2016). Strength model of self-regulation as limited resource: Assessment, controversies, update. *Advances in Experimental Social Psychology*, *54*, 67–127. https://doi.org/10.1016/bs.aesp.2016.04.001

Baumeister, R. F., Wright, B. R., & Carreon, D. (2018). Self-control "in the wild": Experience sampling study of trait and state self-regulation. *Self and Identity*. https://doi.org/10.1080/15298868.2018.1478324

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (in press). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review*. https://doi.org/10.1177/1088868318762183

Gawronski, B., & Bodenhausen, G. V. (2015). Theory evaluation. In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 3–23). New York, NY: Guilford Publications.

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences, 18*, 127–133. https://doi.org/10.1016/j.tics.2013.12.009

Van Lange, P. A. M. (2013). What we should expect from theories in social psychology: Truth, abstraction, progress, and applicability as standards (TAPAS). *Personality and Social Psychology Review, 17*, 40–55. https://doi.org/10.1177/1088868312453088

References

Ampel, B. C., Muraven, M., & McNay, E. C. (2018). Mental Work Requires Physical Energy: Self-Control Is Neither Exception nor Exceptional. *Frontiers in psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01005

Arber, M. M., Ireland, M. J., Feger, R., Marrington, J., Tehan, J., & Tehan, G. (2017). Ego depletion in real-time: An examination of the sequential-task paradigm. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01672

Association for Psychological Science. (2018, July 4). Registered replication reports. Retrieved from https://www.psychologicalscience.org/publications/replication

Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7*, 543–554. https://doi.org/10.1177/1745691612459060

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall. https://doi.org/10.1177/105960117700200317

Baumeister, R. F., Bratlavsky, 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. 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*, 1–15. https://doi.org/10.1207/s15327965pli0701_1

Baumeister, R. F., & Vohs, K. D. (2016). Strength model of self-regulation as limited resource. *Advances in Experimental Social Psychology, 54*, 67–127. https://doi.org/10.1016/bs.aesp.2016.04.001

Beames, J. R., Schofield, T. P., & Denson, T. F. (2017). A meta-analysis of improving self-control with practice. In D. T. D. de Ridder, M. A. Adriaanse, & K. Fujita (Eds.), *Routledge international handbook of self-control in health and well-being* (pp. 405–417). Abingdon, UK: Routledge. https://doi.org/10.4324/9781315648576-32

Blázquez, D., Botella, J., & Suero, M. (2017). The debate on the ego-depletion effect: Evidence from meta-analysis with the p-uniform method. *Frontiers in Psychology, 8*(197). https://doi.org/10.3389/fpsyg.2017.00197

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. Chichester, UK: John Wiley & Sons. https://doi.org/10.1002/9780470743386

Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectations: A test of the general aggression model. *Personality and Social Psychology Bulletin, 28*, 1679–1686. https://doi.org/10.1177/014616702237649

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*, 365–376. https://doi.org/10.1038/nrn3475

Carter, E. C., Kofler, L. M., Forster, D. E., & McCullough, M. E. (2015). A series of meta-analytic tests of the depletion effect: Self-control does not seem to rely on a limited resource. *Journal of Experimental Psychology: General, 144*, 796–815. https://doi.org/10.1037/xge0000083

Carter, E. C., & McCullough, M. E. (2014). Publication bias and the limited strength model of self-control: Has the evidence for ego depletion been overestimated? *Frontiers in Psychology, 5*(823). https://doi.org/10.3389/fpsyg.2014.00823

Carter, E. C., Schönbrodt, F., Gervais, W. M., & Hilgard, J. (2017). *Correcting for bias in psychology: A comparison of meta-analytic methods*. OSF Pre-Print. https://doi.org/10.17605/OSF.IO/9H3NU

Chow, J. T., Hui, C. M., & Lau, S. (2015). A depleted mind feels inefficacious: Ego-depletion reduces self-efficacy to exert further self-control: Ego-depletion and self-efficacy. *European Journal of Social Psychology, 45*, 754–768. https://doi.org/10.1002/ejsp.2120

Christian, M. S., & Ellis, A. P. J. (2011). Examining the effects of sleep deprivation on workplace deviance: A self-regulatory perspective. *Academy of Management Journal, 54*, 913–934. https://doi.org/10.5465/amj.2010.0179

Christiansen, P., Cole, J. C., & Field, M. (2012). Ego depletion increases ad-lib alcohol consumption: Investigating cognitive mediators and moderators. *Experimental and Clinical Psychopharmacology, 20*, 118–128. https://doi.org/10.1037/a0026623

Christley, R. M. (2010). Power and error: Increased risk of false positive results in underpowered studies. *The Open Epidemiology Journal, 3*, 16–19. https://doi.org/10.2174/1874297101003010016

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: L. Erlbaum Associates. https://doi.org/10.4324/9780203771587

Cunningham, M. R., & Baumeister, R. F. (2016). How to make nothing out of something: Analyses of the impact of study sampling and statistical interpretation in misleading meta-analytic conclusions. *Frontiers in Psychology, 7*. https://doi.org/10.3389/fpsyg.2016.01639

Dang, J. (2016). Commentary: A multilab preregistered replication of the ego-depletion effect. *Frontiers in Psychology, 7*(1155). https://doi.org/10.3389/fpsyg.2016.01155

Dang, J., Liu, Y., Liu, X., & Mao, L. (2017). The ego could be depleted, providing initial exertion is depleting. *Social Psychology, 48*, 242–245. https://doi.org/10.1027/1864-9335/a000308

De Houwer, J. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6*, 202–209. https://doi.org/10.1177/1745691611400238

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. https://doi.org/10.1177/1088868313507533

DeWall, C. N., Baumeister, R. F., Stillman, T. F., & Gailliot, M. T. (2007). Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental Social Psychology, 43*, 62–76. https://doi.org/10.1016/j.jesp.2005.12.005

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45*, 259–268. https://doi.org/10.1016/j.jrp.2011.02.004

Egger, M., Smith, G. D., & Sterne, J. A. C. (2001). Uses and abuses of meta-analysis. *Clinical Medicine, 1*, 478–484. https://doi.org/10.7861/clinmedicine.1-6-478

Estrada, M., & Schultz, P. W. (2017). The use of theory in applied social psychology. In A. P. Buunk, K. Keizer, L. Steg, & T. Rothengatter (Eds.), *Applied social psychology: Understanding and managing social problems* (2nd ed., pp. 27–51). Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/9781107358430.002

Etherton, J. L., Osborne, R., Stephenson, K., Grace, M., Jones, C., & De Nadai, A. S. (2018). Bayesian analysis of multimethod ego-depletion studies favours the null hypothesis. *British Journal of Social Psychology, 57*, 367–385. https://doi.org/10.1111/bjso.12236

Evans, D. R., Boggero, I. A., & Segerstrom, S. C. (2015). The nature of self-regulatory fatigue and "ego depletion": Lessons from physical fatigue. *Personality and Social Psychology Review, 20*, 291–310. https://doi.org/10.1177/1088868315597841

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PLoS One, 5*, e10271. https://doi.org/10.1371/journal.pone.0010271

Fanelli, D., Costas, R., & Ioannidis, J. P. (2017). Meta-assessment of bias in science. *Proceedings of the National Academy of Sciences, 114*, 3714–3719. https://doi.org/10.1073/pnas.1618569114

Fiedler, K., & Schwarz, N. (2016). Questionable research practices revisited. *Social Psychological and Personality Science, 7*, 45–52. https://doi.org/10.1177/1948550615612150

Freeman, N., & Muraven, M. (2010). Self-control depletion leads to increased risk taking. *Social Psychological and Personality Science, 1*, 175–181. https://doi.org/10.1177/1948550609360421

Friese, M., Frankenbach, J., Job, V., & Loschelder, D. D. (2017). Does self-control training improve self-control? A meta-analysis. *Perspectives on Psychological Science, 12*, 1077–1099. https://doi.org/10.1177/1745691617697076

Friese, M., Hofmann, W., & Wänke, M. (2008). When impulses take over: Moderated predictive validity of explicit and implicit attitude measures in predicting food choice and consumption behaviour. *British Journal of Social Psychology, 47*, 397–419. https://doi.org/10.1348/014466607x241540

Friese, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review, 23*, 107–131. https://doi.org/10.1177/1088868318762183

Gawronski, B., & Bodenhausen, G. V. (2015). Theory evaluation. In B. Gawronski & G. V. Bodenhausen (Eds.), *Theory and explanation in social psychology* (pp. 3–23). New York, NY: Guilford Publications. Retrieved from http://www.bertramgawronski.com/publications.htm

Graham, J. D., Martin Ginis, K. A., & Bray, S. R. (2017). Exertion of self-control increases fatigue, reduces task self-efficacy, and impairs performance of resistance exercise. *Sport, Exercise, and Performance Psychology, 6*, 70–88. https://doi.org/10.1037/spy0000074

Gröpel, P., Baumeister, R. F., & Beckmann, J. (2014). Action versus state orientation and self-control performance after depletion. *Personality and Social Psychology Bulletin, 40*, 476–487. https://doi.org/10.1177/0146167213516636

Gurevitch, J., Koricheva, J., Nakagawa, S., & Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature, 555*, 175–182. https://doi.org/10.1038/nature25753

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*, 546–573. https://doi.org/10.1177/1745691616652873

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*, 495–525. https://doi.org/10.1037/a0019486

Hartgerink, C. H. J. (2017). Reanalyzing Head et al. (2015): Investigating the robustness of widespread p-hacking. *PeerJ, 5*, e3068. https://doi.org/10.7717/peerj.3068

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. New York, NY: Guilford Press.

Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of p-hacking in science. *PLoS Biology, 13*, e1002106. https://doi.org/10.1371/journal.pbio.1002106

Heatherton, T. F., & Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Sciences, 15*, 132–139. https://doi.org/10.1016/j.tics.2010.12.005

Hempel, C. G. (1970). Aspects of scientific explanation and other essays in the philosophy of science. New York: Free Press.

Hirt, E. R., Clarkson, J. J., Egan, P. M., & Eyink, J. R. (2016). Restoration effects following depletion. In *Self-regulation and ego control* (pp. 227–251). Elsevier. https://doi.org/10.1016/B978-0-12-801850-7.00012-3

Hirt, E. R., Clarkson, J. J., & Jia, L. (2016). *Self-regulation and ego control*. Academic Press. https://doi.org/10.1016/c2014-0-01291-9

Inzlicht, M., Gervais, W., & Berkman, E. (2015). *Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015* (SSRN Scholarly Paper No. ID 2659409). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2659409

Inzlicht, M., & Gutsell, J. N. (2007). Running on empty: Neural signals for self-control failure. *Psychological Science, 18*, 933–937. https://doi.org/10.1111/j.1467-9280.2007.02004.x

Inzlicht, M., Schmeichel, B. J., & Macrae, C. N. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences, 18*, 127–133. https://doi.org/10.1016/j.tics.2013.12.009

Ioannidis, J. P., & Lau, J. (1998). Can quality of clinical trials and meta-analyses be quantified? *The Lancet, 352*, 590. https://doi.org/10.1016/S0140-6736(98)22034-4

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego-depletion - is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science, 21*, 1686–1693. https://doi.org/10.1037/e634112013-142

Job, V., Friese, M., & Bernecker, K. (2015). Effects of practicing self-control on academic performance. *Motivation Science, 1*, 219–232. https://doi.org/10.1037/mot0000024

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*, 524–532. https://doi.org/10.1177/0956797611430953

Kotabe, H. P., & Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science, 10*, 618–638. https://doi.org/10.1177/1745691615593382

Kurzban, R. (2010). Does the brain consume additional glucose during self-control tasks? *Evolutionary Psychology, 8*, 147470491000800200. https://doi.org/10.1177/147470491000800208

Loschelder, D. D., & Friese, M. (2016). Moderators of the ego depletion effect. In E. R. Hirt, J. J. Clarkson, & L. Jia (Eds.), *Self-regulation and ego control* (pp. 21–65). San Diego, CA: Academic Press. https://doi.org/10.1016/b978-0-12-801850-7.00002-0

Luethi, M. S., Friese, M., Binder, J., Boesiger, P., Luechinger, R., & Rasch, B. (2016). Motivational incentives lead to a strong increase in lateral prefrontal activity after self-control exertion. *Social Cognitive and Affective Neuroscience, 11*, 1618–1626. https://doi.org/10.1093/scan/nsw073

Lurquin, J. H., Michaelson, L. E., Barker, J. E., Gustavson, D. E., von Bastian, C. C., Carruth, N. P., & Miyake, A. (2016). No evidence of the ego-depletion effect across task characteristics and individual differences: A pre-registered study. *PLoS One, 11*, e0147770. https://doi.org/10.1371/journal.pone.0147770

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., … Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, 0021. https://doi.org/10.1038/s41562-016-0021

Muraven, M. (2010). Building self-control strength: Practicing self-control leads to improved self-control performance. *Journal of Experimental Social Psychology, 46*, 465–468. https://doi.org/10.1016/j.jesp.2009.12.011

Muraven, M., Collins, R. L., & Neinhaus, K. (2002). Self-control and alcohol restraint: An initial application of the self-control strength model. *Psychology of Addictive Behaviors, 16*, 113–120. https://doi.org/10.1037//0893-164X.16.2.113

Muraven, M., & Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin, 29*, 894–906. https://doi.org/10.1037/e633872013-621

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 201708274*. https://doi.org/10.1073/pnas.1708274114

Osgood, J. M. (2017). Effect of ego-depletion typing task on Stroop does not extend to diverse online sample. *Journal of Articles in Support of the Null Hypothesis, 13*, 83–90. Retrieved from http://www.jasnh.com/pdf/Vol13-No2-article5.pdf

Pocheptsova, A., Amir, O., Dhar, R., & Baumeister, R. F. (2009). Deciding without resources: Resource depletion and choice in context. *Journal of Marketing Research, XLVI*, 344–355. https://doi.org/10.2139/ssrn.955427

Popper, K. R. (1959). *The logic of scientific discovery*. London, UK: Hutchinson.

Savani, K., & Job, V. (2017). Reverse ego-depletion: Acts of self-control can improve subsequent performance in Indian cultural contexts. *Journal of Personality and Social Psychology, 113*, 589–607. https://doi.org/10.1037/pspi0000099

Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: Affirming core values counteracts ego depletion. *Journal of Personality and Social Psychology, 96*, 770–782. https://doi.org/10.1037/a0014635

Schönbrodt, F. D. (2015). Retrieved from http://shinyapps.org/apps/p-hacker/

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366. https://doi.org/10.1177/0956797611417632

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2018). False-positive citations. *Perspectives on Psychological Science, 13*, 255–259. https://doi.org/10.1177/1745691617698146

Singh, R. K., & Göritz, A. S. (2018). Ego depletion does not interfere with working memory performance. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00538

Sjåstad, H., & Baumeister, R. F. (2018). The future and the will: Planning requires self-control, and ego depletion leads to planning aversion. *Journal of Experimental Social Psychology, 76*, 127–141. https://doi.org/10.1016/j.jesp.2018.01.005

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology, 89*, 845. https://doi.org/10.1037/0022-3514.89.6.845

Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science, 25*, 1227–1234. https://doi.org/10.1177/0956797614526415

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less

pathology, better grades, and interpersonal success. *Journal of Personality, 72*, 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Vadillo, M. A., Gold, N., & Osman, M. (2016). The bitter truth about sugar and willpower: The limited evidential value of the glucose model of ego depletion. *Psychological Science, 27*, 1207–1214. https://doi.org/10.1177/0956797616654911

Vadillo, M. A., Gold, N., & Osman, M. (2018). Searching for the bottom of the ego well: Failure to uncover ego depletion in many labs 3. *Royal Society Open Science, 5*, 180390. https://doi.org/10.1098/rsos.180390

van 't Veer, A. E., & Giner-Sorolla, R. (2016). Pre-registration in social psychology—A discussion and suggested template. *Journal of Experimental Social Psychology, 67*, 2–12. https://doi.org/10.1016/j.jesp.2016.03.004

Van Lange, P. A. M. (2013). What we should expect from theories in social psychology: Truth, abstraction, progress, and applicability as standards (TAPAS). *Personality and Social Psychology Review, 17*, 40–55. https://doi.org/10.1177/1088868312453088

Vohs, K. D., Baumeister, R. F., & Schmeichel, B. J. (2012). Motivation, personal beliefs, and limited resources all contribute to self-control. *Journal of Experimental Social Psychology, 48*, 943–947. https://doi.org/10.1016/j.jesp.2012.03.002

Vohs, K. D., Baumeister, R. F., Schmeichel, B. J., Twenge, J. M., Nelson, N. M., & Tice, D. M. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology, 94*, 883–898. https://doi.org/10.1037/0022-3514.94.5.883

Vohs, K. D., & Faber, R. J. (2007). Spent resources: Self-regulatory resource availability affects impulse buying. *Journal of Consumer Research, 33*, 537–547. https://doi.org/10.1086/510228

Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological Science, 11*, 249–254. https://doi.org/10.1111/1467-9280.00250

Whitley, B. E. (2002). *Principles of research in behavioral science* (2nd ed.). Boston, MA: McGraw-Hill. https://doi.org/10.4324/9780203085219

Wolff, W., Baumann, L., & Englert, C. (2018). Self-reports from behind the scenes: Questionable research practices and rates of replication in ego depletion research. *PLoS One, 13*, e0199554. https://doi.org/10.1371/journal.pone.0199554

Yam, K. C., Chen, X.-P., & Reynolds, S. J. (2014). Ego depletion and its paradoxical effects on ethical decision making. *Organizational Behavior and Human Decision Processes, 124*, 204–214. https://doi.org/10.1016/j.obhdp.2014.03.008