Investigating Ego Depletion Phenomena from Diverse Angles

Dissertation

zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

der Fakultät HW

Bereich Empirische Humanwissenschaften

der Universität des Saarlandes

vorgelegt von

Karolin Ulrike Gieseler

aus Bonn

Saarbrücken, 2020

Dekan:	Univ.-Prof. Dr. Stefan Strohmeier, Universität des Saarlandes
Berichterstatter:	Univ.-Prof. Dr. Malte Friese, Universität des Saarlandes
	Univ.-Prof. Dr. Cornelius J. König, Universität des Saarlandes
Tag der Disputation:	17.12.2019

Summary

Self-control is a fundamental capacity in life. It enables people to act in accordance with their short- and long-term goals and to inhibit undesired automatic action tendencies. However, people are not always successful in exerting self-control. Researchers typically investigate self-control failure in the laboratory. They predominantly use the *sequential task paradigm*: The first of two consecutive tasks manipulates self-control demands (*high* versus *low*) while the second task assesses self-control performance as a function previous self-control exertion. If participants show impaired self-control performance after facing *high* compared to *low* self-control demands in the first task, an *ego depletion* effect is observed. Several theoretical accounts propose, among others, a limited self-control resource or diminished motivation to explain ego depletion phenomena. Even if hundreds of studies provided support for the ego depletion idea, recent meta-analyses and replication attempts raised substantiated doubts—mainly due to publication bias and *p*-hacking—as to the very existence of the phenomenon. As the ego depletion idea has grown very influential in research as well as in public perception, it seems premature to abandon the idea altogether as long as the existing evidence is inconclusive. The present thesis addresses several shortcomings of the ego depletion literature and tries to contribute to the methodological and theoretical progress of the field.

As a first step, two studies investigated the feasibility to examine ego depletion effects within participants in one experimental session with increased statistical power. The results indicated that ego depletion effects are not easy to examine within participants using the sequential task paradigm twice in a row. This was partly due to strong learning effects and difficulties to re-establish pre-study conditions for the second sequential task paradigm. However, the studies provided pre-registered support for the ego depletion effect between participants as a by-product of the counterbalanced design.

As a second step, my colleagues and I turned to possible underlying mechanisms proposed by the two prominent models in ego depletion research. We therefore extended the traditional sequential task paradigm by two additional components: the established moderator self-affirmation and psychophysiological indicators of mental effort. This way, we created a context that allows to contrast the conflicting predictions of the two models. Due to the unexpected effect of the moderator self-affirmation, we were limited in our analysis contrasting the models. Nevertheless, the study shed light on possible effects of self-affirmation and the progression of mental effort over the course of a study as a promising subject of future research.

As a last step, three studies used an alternative dependent measure to assess the effects of *high* versus *low* self-control demands: mental effort choice. We reasoned that in daily life, people are often free to influence their course of action and the amount of effort they want to invest. The heterogeneous findings emphasize the importance of self-rated ability and self-efficacy, and the relevance of the subjective experience of effort.

In sum, my thesis suggests that it would be rash to abandon the ego depletion idea altogether and that mental effort choice might be a promising alternative to the traditional dependent measures. Psychophysiological indicators of mental effort seem important to shed light on the underlying mechanisms—as are subjective experiences of mental effort. The general discussion points to the maybe most important place for future research: daily life.

Zusammenfassung

Selbstkontrolle ist eine grundlegende Fähigkeit im Leben. Sie ermöglicht es uns, in Übereinstimmung mit unseren kurz- und langfristigen Zielen zu handeln und unerwünschte automatische Handlungstendenzen zu unterdrücken. Menschen sind jedoch nicht immer erfolgreich in der Ausübung von Selbstkontrolle. Wissenschaftler untersuchen das Versagen von Selbstkontrolle typischerweise im Labor untersucht. Dazu verwenden sie überwiegend das *sequentielle Task Paradigma*: Die erste von zwei aufeinanderfolgenden Aufgaben manipuliert die Selbstkontrollanforderungen (*hoch* versus *niedrig*) während die zweite Aufgabe Selbstkontrollperformanz als Funktion vorheriger Selbstkontrollausübung misst. Wenn die Versuchsteilnehmer verminderte Selbstkontrollperformanz zeigen, nachdem sie *hohen* im Vergleich zu *niedrigen* Selbstkontrollanforderungen in der ersten Aufgabe ausgesetzt waren, spricht man von einem *Ego Depletion Effekt*. Verschiedene theoretische Ansätze schlagen, unter anderem, eine begrenzte Selbstkontroll-Ressource oder verringerte Motivation als Erklärung für Ego Depletion Phänomene vor. Obwohl hunderte Studien die Ego Depletion Idee empirisch untermauern, wecken jüngere Meta-Analysen und Replikationsversuche erhebliche Zweifel an der bloßen Existenz des Phänomens, vor allem aufgrund von Publication Bias und *P*-Hacking. Da die Ego Depletion Idee in der Forschung sowie in der öffentlichen Wahrnehmung sehr einflussreich geworden ist, erscheint es voreilig, die Idee als solche aufzugeben, solange die bestehende Evidenz nicht eindeutig für oder gegen sie spricht. Die vorliegende Dissertation adressiert verschiedene Defizite der Ego Depletion Literatur und versucht zum methodischen und theoretischen Fortschritt des Feldes beizutragen. In einem ersten Schritt untersuchen zwei Studien die Möglichkeit Ego Depletion Effekte innerhalb von Personen in einer experimentellen Sitzung mit erhöhter statistischer Power. Die Ergebnisse deuteten darauf hin, dass Ego Depletion Effekte nicht uneingeschränkt innerhalb von Personen untersucht werden können, indem man

das sequentielle Task Paradigma zweimal in Folge verwendet. Das liegt unter anderem in starken Lerneffekten begründet sowie in Schwierigkeiten, die Versuchspersonen für den zweiten Teil der Studie in einen Zustand zurückzuversetzen, der dem zu Beginn der Studie gleicht. Die Studien liefern jedoch prä-registrierte Evidenz für den Ego Depletion Effekt zwischen Personen als Nebenprodukt des counterbalancierten Designs. Anschließend wandten meine Kollegen und ich uns möglichen zugrundeliegenden Mechanismen zu, die die zwei prominentesten Modelle der Ego Depletion Forschung vorschlagen. Zu diesem Zweck haben wir das traditionelle sequentielle Task Paradigma um zwei Komponenten erweitert: den Moderator Self-Affirmation und psychophysiologische Indikatoren mentaler Anstrengung. Auf diese Weise schufen wir einen Kontext, der es ermöglicht die widersprüchlichen Vorhersagen der Modelle zu kontrastieren. Aufgrund des unerwarteten Effekts des Moderators Self-Affirmation waren wir in der Analyse, die die Modelle einander gegenüberstellt, eingeschränkt. Dennoch warf die Studie Licht auf mögliche Effekte von Self-Affirmation und die Entwicklung mentaler Anstrengung im Verlauf einer Studie als vielversprechende Wege zukünftiger Forschung. In einem letzten Schritt nutzten drei Studien eine alternative abhängige Variable um die Effekte *hoher* gegenüber *niedriger* Selbstkontrollanforderungen zu messen: die Wahl mentaler Anstrengung. Unserer Überlegung nach sind Menschen im täglichen Leben oft frei in der Gestaltung ihres Vorgehens und dem Ausmaß mentaler Anstrengung, das sie zu investieren bereit sind. Die heterogenen Befunde unterstreichen die Bedeutsamkeit selbstbeurteilter Fähigkeit und Selbstwirksamkeit und die Relevanz subjektiv erlebter Anstrengung. Zusammengenommen suggeriert meine Dissertation, dass es verfrüht wäre, die Ego Depletion Idee nicht weiter zu verfolgen und dass die Wahl mentaler Anstrengung eine vielversprechende Alternative zu traditionellen abhängigen Maßen darstellt. Psychophysiologische Indikatoren mentaler Anstrengung scheinen wichtig um Licht auf zugrundeliegende Mechanismen zu werfen, wie auch die subjektive Wahrnehmung mentalen

Aufwands. Die abschließende Diskussion verweist auf den vielleicht wichtigsten Ort zukünftiger Forschung: den Alltag.

Acknowledgements

I would never have started let alone realized this project if it wasn't for the invaluable support of many significant people. I am especially grateful for the unfailing support of my supervisor Malte Friese, for his open door, his appreciative feedback, his expertise and precious advice. I thank Veronika Job and David D. Loschelder for their encouragement and fruitful collaboration on our joint project. I am grateful to Cornelius J. König for agreeing to review my thesis.

My years at university would not have been have as livable, productive and memorable without my colleagues. Thank you, Marcel, Dorota, Janet, Hannah and Lasse for the all lunch and coffee breaks, for scientific and private talks and for your help and cheering up. Special thanks to Annika, Diana, Nicolas and Michelle for collecting countless research participants and to Helge for his help in the physio-lab. For practical advice during the last weeks, thanks to Markus for countless mails.

Last but not least, I want to thank my family and friends for their unconditional support. Carolin, thank you for sharing doubts, concerns, and fears but also successes, progress and many conferences with me. Steffi, thanks for endless telephone calls—even if I will always prefer your kitchen. Special props to Hanna and Bene—always yours and shredding the balloon. Andi and Doro, thank you for always believing in me. Philipp, even if it's not about numbers, I would neither have conquered 7b+ nor written any amount of words if it wasn't for you.

Index of Publications

This publication-oriented dissertation[1] (German: publikations-orientierte Dissertation) is mainly based on three manuscripts, two of them currently submitted for publication, one of them invited for resubmission. The author of this dissertation is the first author of all of these manuscripts which are inserted in this dissertation in their most recent version (slight changes due to formatting). All of the studies were pre-registered on the Open Science Framework (osf.io).

Part I: Gieseler, K., Loschelder, D. D., Job, V., Herrmann, C., Friese, M. *Pre-registered evidence for the ego depletion effect between, but not within persons.* Manuscript invited for resubmission to the Journal of Experimental Social Psychology.

Part II: Gieseler, K., Inzlicht, M., Friese, M. (2019). *Do people avoid mental effort after facing a highly demanding task?* Manuscript submitted for publication to the Journal of Experimental Social Psychology.

Part III: Gieseler, K., Loschelder, D. D., Job, V., Friese, M. *A pre-registered test of competing theories to explain ego depletion effects using psychophysiological indicators of mental effort.* Manuscript submitted for publication to Motivation Science.

---

[1] The layout of this dissertation follows the model of Langer (2018) and Miketta (2018).

Content

Introduction

Preliminary remark: The introduction of this thesis overlaps to a some extent with a book chapter I first-authored on the quality and evaluation of theories and accompanying empirical research using a prominent model in self-control research as illustrative reference point (Gieseler, Loschelder, & Friese, 2019) and with a meta-science paper I contributed to analyzing the current state of the research field of self-control depletion (Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2019). In my thesis, I therefore put more emphasis on the General Discussion than on the Introduction regarding my own reflections and ideas.

"I count him braver who overcomes his desires than him who conquers his enemies; for the hardest victory is over self." – Aristotle

Control of the Self has long fascinated, troubled and despaired mankind. Going to bed early, doing sports, eating healthy and living in harmony with our beloved ones are only some of many goals we pursue in life. However, staying up late to binge-watch a thrilling series while indulging potato chips, spreading on the couch because tomorrow is another day for training and venting one's anger if the kids as usual did not close the toothpaste tube are common behaviors disregarding long-term goals. Most people would agree that if it was easier for them to not give in to temptations, to always pursue their long-term goals, to control their emotions, thoughts and behaviors, they would have fewer problems in life. From a broader perspective, if individuals succeeded to do so, this would be beneficial for societies spending millions to counteract the adverse consequences of an unhealthy lifestyle, dealing with high crime rate, poverty and ultimately even climate change and environmental pollution caused by thoughtlessness and short-term gratification.

Psychological research invested many resources studying the ability supposedly enabling people to reach long-term goals: self-control. In the laboratory, researchers used a so called sequential-task paradigm: People exert self-control in a first task and their self-control performance in the subsequent task is compared to participants who did not control themselves in the initial task. Over more than a decade, a large literature emerged studying self-control failure in the laboratory. Because of its relevance and immediate comprehensibility, research on self-control and its failure also gained public awareness. In recent years however, doubts arose as to the robustness and viability of the research and the paradigm predominantly used. Many researchers question the meaningfulness of the results, the research methods and the literature on self-control failures in general.

This thesis explores several possibilities to contribute to the literature on self-control failure in the present circumstances characterized by doubts and critique. First, the introduction provides an overview of self-control research and ego depletion phenomena, two important theoretical accounts, recent developments and the present research (Part I, II and III). In the following, Part I tests the possibility to study self-control failures in the laboratory using within-participants designs to increase statistical power. Part II focuses on the underlying mechanisms addressing the facet of insufficient theorizing. Part III expands theoretical considerations using a new form of measure for effects of self-control depletion with a focus on choice instead of performance. The general discussion summarizes the findings of the present research, points to limitation and provides reflections on psychological research, its relationship to common sense and future directions.

Self-Control and Ego Depletion Phenomena

In the collective memory of psychological researchers, the most recent wave of research on self-control started with Mischel's famous "marshmallow experiments" (Mischel & Ebbesen, 1970; Mischel, Shoda, & Rodriguez, 1989). The general procedure was as follows: He left preschoolers alone in a room with nothing but a single marshmallow. They were told that if they awaited the experimenter's return without touching the candy, they would be rewarded a second one. This ability to resist the immediate gratification in favor of a greater good has seen many names over the last decades. Willpower (Job, Dweck, & Walton, 2010), self-regulation (Baumeister & Heatherton, 1996), self-control (Tangney, Baumeister, & Boone, 2004), Ego strength (Kahan, Polivy, & Herman, 2003), executive (Engle, 2002) or inhibitory control (Engle, Conway, Tuholski, & Shisler, 1995), delay of gratification (e.g. Mischel & Ebbesen, 1970), behavioral restraint (Wright, Mlynski, & Carbajal, 2019) or the antonym impulsivity (Duckworth & Kern, 2011) all refer to "the 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 et al., 2004, p. 274) or a related idea. Because of its broad usage, self-control is sometimes referred to as "an umbrella construct that bridges concepts and measurements from different disciplines" (Moffitt et al., 2011, p. 2693). In the present thesis, I will use the term "self-control" along with Tangney's definition.

In 1994, Baumeister and colleagues published the results of their extensive literature review on self-control research (Baumeister, Heatherton, & Tice, 1994). They observed that people were more likely to fail at controlling themselves after exerting self-control beforehand. Especially after experiencing stress or after a long and tiring day, people were more prone to self-control failures such as breaking their diets, committing crimes, or consuming alcohol and cigarettes (Baumeister & Heatherton, 1996). To describe the phenomenon that the initial exertion of self-control makes subsequent failures in self-control more probable, Baumeister and colleagues introduced the term of "ego depletion" (Baumeister, Bratlavsky, Muraven, & Tice, 1998). Ego depletion thus refers to the observation that on the functional level of analysis in terms of observable behavior (De Houwer, 2011), exerting self-control in one domain leads to decreased self-control performance in the same or any other self-control domain. The term *ego depletion* is limited to behavioral effects and does not relate to the underlying processes and possible mechanisms accounting for the phenomenon on the cognitive level of analysis (De Houwer, 2011).

In the laboratory, ego depletion phenomena are since predominantly tested using a sequential-task paradigm: Participants work on two consecutive tasks with the first being either highly self-control demanding or a control task with low self-control demands. The second task assesses self-control performance in function of the self-control demands of the first task (Baumeister et al., 1998; Muraven, Tice, & Baumeister, 1998). Over about two decades starting in the mid-nineties of the 20th century, hundreds of studies used this paradigm to test the ego depletion idea: Participants exerting self-control in the initial task should show worse self-control performance in the second task compared to those initially working on a control task not requiring self-control (for meta-analyses, see Carter, Kofler, Forster, & McCullough, 2015; Dang, 2018; Dang, Björklund, & Bäckström, 2017; Hagger, Wood, Stiff, & Chatzisarantis, 2010). For instance, after suppressing their thoughts or emotions or after resisting tempting food, participants consumed more alcohol in a subsequent product test (Muraven, Collins, & Nienhaus, 2002), lost stamina faster in a hand-grip task (Muraven et al., 1998) and persisted shorter in an unsolvable puzzle task (Baumeister et al., 1998).

Theoretical Accounts for Ego Depletion Phenomena

There are several theoretical accounts to explain the underlying mechanisms of ego depletion phenomena. The most prominent among them is the *strength model of self-control* (Baumeister et al., 1998; Baumeister & Heatherton, 1996; Muraven et al., 1998). It was developed along with the observations of ego depletion effects in everyday life and the initial review of the literature by Baumeister and colleagues (Baumeister et al., 1994). They concluded that self-control is a resource that is characterized by being domain-general and limited. The analogy of a muscle is often used to illustrate the strength model and its assumptions: A muscle gets tired when used regardless if it is used for running or cycling. A resource of self-control is used up when spend regardless if used to control the impulse to smoke or to indulge sweets, to abandon the couch to train or to inhibit the automatic reactions to incongruent Stroop trials. However, a muscle can also be trained and Baumeister and colleagues reasoned that extensive use would increase the capability to exert self-control over time (for meta-analyses, see Beames, Schofield, & Denson, 2017; Friese, Frankenbach, Job, & Loschelder, 2017). Thus, on the cognitive level of analysis (De Houwer, 2011), the strength model assumes the reduction of a resource as the explanation for ego depletion effects in terms of behavior on the functional level of analysis. Although a lot of researchers invested in the discovery of the nature of this resource, it remains elusive. The most tangible suggestion was glucose as a limited energy resource, but the results are inconclusive (Ampel, Muraven, & McNay, 2018; Gailliot et al., 2007; Vadillo, Gold, & Osman, 2016).

The *process model of self-control*—one of the most prominent alternative models—takes a different approach (Inzlicht & Schmeichel, 2012). It does not assume a limited resource but motivational shifts to underlie ego depletion effects: Engaging in demanding tasks reduces the motivation to exert further control and leads to a shift from so called “have-to” to “want-to” goals. According to the model, people are still *able* to exert self-control, but they are no longer motivated to do so and strive for easier and more agreeable pastime. Along with these motivational processes attention and emotions vary accordingly. Ego depletion effects occur if “have-to” goals, that often require actively resisting behavioral impulses, get out of focus in favor of goals aiming at self-gratification. Part II of this thesis is dedicated to the differing predictions of the strength and the process model and proposes a possibility to contrast their assumptions.

Recent Developments

Since the first literature review and proposition of the strength model (Baumeister & Heatherton, 1996), ego depletion phenomena were extensively studied. A first meta-analysis (Hagger et al., 2010) attested the ego depletion effect a solid mean effect size of $d = 0.62$. However, this meta-analysis did not include unpublished studies and the medium-to-large effect size was later suggested to be inflated due to publication bias in a re-analysis, raising first doubts if the ego depletion effect is distinguishable from zero (Carter & McCullough, 2014). A second meta-analysis addressed some of the shortcomings of the first and concluded that the evidence does not strongly suggest that the effect exists at all (Carter et al., 2015). Additional doubts on the very existence of ego depletion effects were raised when a large-scale, high-powered registered replication report (RRR) revealed an overall null effect over 23 laboratories (Hagger et al., 2016). More recently, several studies found further null effects – some of them following pre-registration protocols (Etherton et al., 2018; Koppel, Andersson, Västfjäll, & Tinghög, 2019; Lurquin et al., 2016; Osgood, 2017). Pre-registered evidence in favor of ego depletion phenomena remains scarce (Dang, Liu, Liu, & Mao, 2017; Garrison, Finley, & Schmeichel, 2019).

Limitations apart from meta-analyses and RRR concern the empirical evidence in general. Manipulation checks pertaining to the successful manipulation of self-control demands in laboratory research were often weak or not employed at all (e.g., see Hagger et al., 2010). It is thus often impossible to say if a certain study tested the postulated effect. In many cases, crucial information as to whether participants invested more in the condition high in self-control demands is missing. Furthermore, the variety of tasks used to evoke ego depletion is very large—as is the range of dependent measures assessing self-control performance. Whether they validly and reliably induce and measure ego depletion effects is hard to say and seldom if ever addressed in the ego depletion literature. One last limitation pertains to reverse depletion effects. If—overall—the ego depletion effect was indistinguishable from zero, there should be a large amount of studies showing reverse ego depletion effects. However, there are hardly any (Converse & DeShon, 2009; Tuk, Zhang, & Sweldens, 2015; Van Reet, 2015). This could either be explained by a gigantesque file drawer of several thousand studies showing reverse effects (for simulations, see Friese et al., 2019) or be interpreted as indirect evidence in favor of ego depletion effects.

Overall, the evidence regarding the existence of ego depletion phenomena remains inconclusive; neither is their strong evidence that there is something to the phenomenon, nor does the existing evidence allow to rule out this possibility (for a review and detailed analysis, see Friese et al., 2019). As the phenomenon has gained prominence over the last years and as it has important implications for societies—at least because they believe in its existence—it does not seem wise to abandon it altogether. Undoubtedly however, research on ego depletion effects needs to improve in terms of empiricisms, methods and theory development.

The Present Research

The present research addresses several shortcomings of the ego depletion literature discussed in the previous section and aims at contributing to the advancement of the research field in terms of methods and theory development. As a first step, my colleagues and I investigated the ego depletion idea within participants in the laboratory using the traditional dual task paradigm twice, separated by a recreational period. This was an attempt to increase statistical power without increasing sample size—resources in psychological research are scarce and need to be used wisely. Knowing if and under which circumstances it is possible to use the sequential task paradigm within participants would be beneficial when further investigating ego depletion phenomena. As a second step, we turned to the underlying mechanisms proposed by the strength and the process model. As both models make the same predictions on the functional level of analysis in terms of observable behavior, we combined several components to create a context that allows to contrast the different predictions on the cognitive level of analysis: the traditional sequential-task paradigm, a moderator and multiple psychophysiological indicators of mental effort. As a last step, we aimed at extending knowledge on ego depletion effects on the functional and the cognitive level of analysis using a different indicator of self-control depletion. Reflecting on ego depletion phenomena observed in everyday life, we reasoned that often it is not only crucial *how well* people perform in terms of self-control, but it may be equally crucial if they *choose* the more or less demanding course of action after exerting self-control in the first place.

All studies included in this thesis were pre-registered on the Open Science Framework (osf.io/uyda7/)—most of them before data collection but all of them before processing any data in any sort. We thereby addressed the called for open science due to the well-founded and substantial concerns and critique regarding the ego depletion literature with respect to researchers' degrees of freedom as *p*-hacking and HARKing.

The order of the studies presented does only partly follow chronological order and is based on theoretical considerations.

Part I: Pre-Registered Evidence for the Ego Depletion Effect Between, but not Within Persons

The existing evidence on ego depletion effects to date is best described as inconclusive. One crucial insight of the ongoing debate is that *if* there is something to the ego depletion phenomenon, possible effects are likely of small to moderate size. Effects of small-to-moderate size need appropriate, large numbers of participants to be detected with adequate statistical power. As of yet however, sample sizes in ego depletion research were often insufficiently large and statistical power is chronically low (Carter et al., 2015; Carter & McCullough, 2014). To increase statistical power without increasing sample size, repeated-measures, within-persons designs can be used. As correlations between repeated dependent measures are typically positive in psychology, person-related variance can be accounted for which increases statistical power for a given sample size. The first part thus constitutes a pre-registered attempt to implement and empirically test a research design manipulating self-control demands (*high* versus *low*) within persons in one experimental session only. Recently published research has implemented a within-subject, repeated measures ego depletion paradigm—however, the structure used contained multiple blocks (between 6 and 22 blocks) of very short intervals of self-control demands manipulation or recovery phases (Francis, Milyavskaya, Lin, & Inzlicht, 2018); the procedure thus differed substantially from the traditional sequential task paradigm with regard to the time courses. The design introduced in Part I was similar to previous sequential task paradigms in terms of task lengths and consisted of two blocks, each comprising the self-control demands manipulation (*high* versus *low*) and the dependent measure. The two blocks were separated by a recreational period lasting several minutes to re-establish pre-study conditions. As half of the participants started with the either the *high* or the *low* self-control demands condition, the design also allowed to investigate the traditional between-participants ego depletion effect—albeit with reduced power.

Part II: A Pre-Registered Test of Competing Theories to Explain Ego Depletion Effects Using Psychophysiological Indicators of Mental Effort

There are several possible explanations for ego depletion phenomena on the cognitive level of analysis in terms of underlying mechanisms. Two prominent theoretical models—the strength model (Baumeister & Vohs, 2016) and the process model (Inzlicht & Schmeichel, 2012; Inzlicht, Schmeichel, & Macrae, 2014)—propose fundamentally different explanations for ego depletion effects. However, they make identical predictions for ego depletion effects in terms of observable behavior on the functional level of analysis. Observing self-control performance thus does not suffice to contrast the predictions of the two models. The presented study extends the sequential task paradigm by a moderator previously shown to counteract ego depletion effects (self-affirmation, see Schmeichel & Vohs, 2009) and makes additional use of psychophysiological indicators of mental effort (Richter, Gendolla, & Wright, 2016; Wright et al., 2019). The combination of these three components—the sequential task paradigm assessing self-control performance, the moderator self-affirmation, and psychophysiological measures of mental effort—creates a context that in principle allows to provide empirical evidence that favors one model or the other.

Part III: Do People Avoid Mental Effort After Facing a Highly Demanding Task?

The last part of this thesis takes a different approach to measure the effects of the manipulation of self-control demands. My colleagues and I reasoned that in daily life people are often free to choose between several courses of action differing in the amount of mental effort demanded. As in Part I and Part II, the sequential task paradigm was used as a basis. However, instead of assessing participants' self-control performance in a second task with identical difficulty for everybody, we investigated whether participants prefer less effort-demanding tasks when given the choice. We hypothesized that participants would select options requiring less mental effort in the second task after facing *high* (as opposed to *low*) self-control demands in the first task.

PART I

Pre-Registered Evidence for the Ego Depletion Effect Between, but not Within Persons

Karolin Gieseler
Saarland University

David D. Loschelder
Leuphana University of Lueneburg

Veronika Job
Technical University Dresden

Christina Herrmann
Saarland University

Malte Friese
Saarland University

Manuscript invited for resubmission to the *Journal of Experimental Social Psychology*.

Abstract

Ego depletion—impaired self-control performance after a first demanding task—has been a popular research topic for two decades but is now heavily questioned. Many researchers believe that, if the effect exists at all, it is rather small. Small effects are often unfeasible to investigate with adequate power in between-persons designs, because they require large sample sizes. By contrast, within-persons designs allow for greater power and therefore require smaller sample sizes. Here, we examined the feasibility to investigate ego depletion within persons in a single experimental session in two pre-registered studies with open data and open materials (OSF; osf.io/k287r/; osf.io/my5x2/). Participants worked on two experimental sections, each consisting of a counting task (self-control demands [SCD] *high* / *low*) followed by a Stroop task (dependent variable), respectively. Task order (*high SCD first* / *low SCD first*) was counterbalanced across participants. A recreational period between the two sections was intended to restore pre-study conditions. This design also allowed for examining the traditional between-persons ego depletion effect based on the first section of the design, albeit with reduced statistical power. Results revealed no reliable evidence for ego depletion *within* persons. However, the traditional ego depletion effect *between* persons was significant in both studies. We discuss difficulties of manipulating ego depletion within persons and implications of the present findings for ego depletion research.

Keywords: self-control, ego depletion, within design, pre-registration, Stroop task.

Pre-Registered Evidence for the Ego Depletion Effect Between, but not Within Persons

Ego depletion refers to the effect of impaired self-control performance after previous self-control exertion (Baumeister & Vohs, 2007). This effect inspired an abundance of research across various sub-disciplines of psychology. Recently, however, the existence of the effect has been heavily questioned. Many researchers believe that—if the ego depletion effect exists at all—it is likely of small to moderate size. We present the results of two pre-registered studies that examined ego depletion in within-persons designs that allow for greater statistical power and may thus be more feasible to detect small-to-moderate effects than between-persons designs.

Self-Control and the Ego Depletion Effect

Self-control is the ability to control dominant responses such as impulses, action tendencies, or emotions (Tangney, Baumeister, & Boone, 2004). The ego-depletion hypothesis suggests that when people exert self-control in an initial task, they are more susceptible to poorer performance in a subsequent self-control task (Baumeister & Vohs, 2007). Several hundred studies have supported this hypothesis. For instance, people who suppressed their thoughts or emotions or inhibited automatic responses subsequently consumed more alcohol or unhealthy food, and showed impaired executive control and persistence compared to control groups (Friese, Binder, Luechinger, Boesiger, & Rasch, 2013; Graham & Bray, 2015; Johns, Inzlicht, & Schmader, 2008; Muraven, Collins, & Nienhaus, 2002).

In 2010, a first meta-analysis attested the ego depletion effect a medium-to-large effect size of $d = 0.62$ (Hagger, Wood, Stiff, & Chatzisarantis, 2010). However, a re-analysis of the same data set and a new meta-analysis comprising mostly studies not included in the initial data set suggested that the field was plagued by publication bias—the tendency of authors, reviewers, and journals to preferentially publish significant results (Carter, Kofler, Forster, & McCullough, 2015; Carter & McCullough, 2014). Carter and colleagues found smaller and, at times, non-significant effect sizes and concluded that the ego depletion effect may be indistinguishable from zero (Carter et al., 2015: corrected mean effect size estimates [trim-and-fill, PET, PEESE] ranging from $g = -0.27$ to $g = 0.24$; Carter & McCullough, 2014: bias-corrected mean effect size estimates ranging from $g = -0.10$ to $d = 0.50$). Doubts were further substantiated by a large-scale registered replication report (RRR) realizing one specific combination of an ego depletion manipulation and a dependent variable (Hagger et al., 2016). Across 23 laboratories the average effect of this RRR was close to zero and not significant.

The discrepancy between the seemingly overwhelming evidence for ego depletion (e.g., Baumeister & Vohs, 2016) and the recent, sobering concerns that ego depletion may not be real eventually, has sparked intensive debates in scientific journals and the social media (Blázquez, Botella, & Suero, 2017; Dang, 2016; Hagger & Chatzisarantis, 2014; Inzlicht, Gervais, & Berkman, 2015; see Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2019, for a review). While a review of this debate is beyond the scope of the present article, one insight is particularly important: If ego depletion is a real phenomenon, the effect size is likely smaller than initially expected based on the Hagger et al. (2010) meta-analysis. To reliably detect small-to-moderate effect sizes with high statistical power, researchers need large numbers of participants. Sample sizes in past ego depletion research, though, are often small and statistical power is generally low (Carter et al., 2015; Carter & McCullough, 2014). One strategy to increase statistical power is the use of within-persons designs. In these designs, repeated measurements of the dependent variable typically correlate, allowing researchers to account for person-related variance, which in turn increases statistical power.

The Present Research

The aim of the present research was to examine a novel approach to manipulate ego depletion *within* persons in a single experimental session: One half of participants starts with a demanding task before working on the dependent variable for the first time. After a recreational period intended to re-establish pre-study conditions, participants continue with a non-demanding control task before working on the dependent variable for the second time. This order of tasks is reversed for the other half of participants. We expected participants to perform worse after the demanding as compared to the non-demanding task, irrespective of whether they started with the demanding or the non-demanding task. We conducted two studies following this design.

Note that because both conditions start with either the high-demanding or the low-demanding task first, the experimental design also allows for investigating the traditional between-persons ego depletion effect based on the data of the first section of the study (albeit with reduced power, because we conducted our a-priori power analyses for the focal within-persons effect). The traditional between-persons ego depletion effect was not a focal goal of the present research. Nevertheless, the study design provided the opportunity to examine ego depletion between persons under pre-registered conditions as an ancillary aim.

Both studies were pre-registered on the Open Science Framework (OSF; osf.io/k287r/; osf.io/my5x2/). All materials, data, and analysis scripts can be found in the associated OSF projects (osf.io/8zs93/; osf.io/t6v4a/). These two studies are the only studies that we conducted in this project, precluding the existence of publication bias. To keep the present manuscript succinct, we report some pre-registered parts of the Method and Results sections in the supplemental material. In addition to frequentist statistics, we report Bayes Factors that contrast the fit of the observed data under the null versus under the alternative hypothesis (Jarosz & Wiley, 2014).

Study 1

Method

Participants and design. Ninety-two students (M_{age} = 23.92, *SD* = 3.55; 69 females; no psychology students) were allocated to a 2 (Self-control demands [SCD]: *high* vs. *low*) × 2 (Task order: *high SCD first* vs. *low SCD first*) mixed design with the first factor varied within persons. Stroop interference based on error rates was the main dependent variable.

To our knowledge, there is no prior work on the within-persons ego depletion effect that we could base our estimate on. Hence, we opted for $f = 0.175$ ($d = 0.35$)—a small-to-medium effect according to common conventions (Cohen, 1988). An a-priori power analysis using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) revealed a minimum sample size of 42 participants to achieve a statistical power of $1\text{-}\beta = .80$ for the within-persons effect, assuming a correlation of $r = .70$ of the dependent variable across measurement occasions and $\alpha = 0.05$.

After collecting data from 48 participants, we checked the empirical correlation of the dependent variable across measurement occasions and found that it was substantially smaller than expected ($r = .40$, $p = .005$). We therefore recalculated the minimum required sample size based on the empirical correlation and continued data collection. In the results section, we report $p_{augmented}$ (Sagarin, Ambler, & Lee, 2014) for our main focal effect, the ego depletion effect within-persons. $P_{augmented}$ estimates the type-1 error inflation based on initial and additional sample size, and critical significance values to transparently disclose the impact of peeking at the data.

Procedure. Participants were assigned to one of five cubicles, filled out an informed consent and put on noise-cancelling headphones. They read the general instructions and agreed to participate conscientiously. After filling in a questionnaire assessing mental fatigue, they worked on either the difficult (high SCD) or easy (low SCD) version of the initial task followed by a manipulation check and again mental fatigue questions. Participants then engaged in the first Stroop task and subsequently answered questions assessing their motivation and perceived difficulty of the Stroop task. During a recreational period, separating the first from the second section of the study, all participants watched two film clips (total length: 8:55 min). The goal of this recreational period was for participants to relax and to re-establish pre-study conditions. The second experimental section followed the same structure as the first. The initial task, however, was now presented in the version not worked on before (SCD: *high* vs. *low*). At the end of the study, participants filled in demographic and control questions. Finally, participants were debriefed, paid, and thanked[1].

Manipulation of self-control demands. To manipulate self-control demands, we used a counting task (Kirschbaum, Pirke, & Hellhammer, 1993). In the *high SCD* condition, participants were asked to count backwards from 4,022 in steps of 13 as fast as possible. They were instructed to restart from the beginning whenever they made a mistake. In the *low SCD* condition, participants were asked to count slowly, in a comfortable pace upwards from zero in steps of two. They were instructed to simply continue counting if they were to make a mistake. Participants spoke directly into a microphone placed on the table until a 'Stop' sign appeared on the screen after 6:30 min. Difficult arithmetic operations have previously been used to examine ego depletion (Converse & Deshon, 2009).

Recreational period. During the recreational period, participants watched two film clips. Previous research suggests that a positive mood and relaxation can counteract ego depletion (Tice, Baumeister, Shmueli, & Muraven, 2007; Tyler & Burns, 2008). Therefore, the first video (4:45 min) was an excerpt from a humorous interview with the tennis player

[1] This is an abbreviated description of the procedure. For a complete list of assessed variables, see supplement.

Roger Federer. The second video (4:10 min) was a relaxing excerpt from the documentary "Wild Faces of the Andes" (Sailer, 2011).

Measures.

Stroop Task. The words "red", "green", "blue", and "yellow" were presented on a black screen. When a word appeared, participants had to indicate its font color by pressing one of four keys while ignoring the semantic meaning of the word. On congruent trials, the font color matched the semantic meaning of the word (e.g., "red" written in red font). On incongruent trials, there was a mismatch between the font color and the semantic meaning (e.g., "red" written in green font). The Stroop task requires self-control because on incongruent trials participants need to inhibit the dominant response of pressing the key matching the semantic meaning instead of the font color.

In the first experimental section, at the beginning of the Stroop task, participants completed 18 practice trials with error feedback. The successive Stroop task consisted of 180 trials (120 congruent, 60 incongruent; no feedback). Trials were presented in a pseudorandomized fixed order. Based on previous studies using a similar Stroop task (Friese et al., 2013; Luethi et al., 2016), Stroop interference based on errors (relative number of errors on incongruent minus congruent trials) served as the main dependent variable. As reaction times in the Stroop task are also a common dependent variable, we included them as well.

Mental fatigue. Mental fatigue was assessed with four items from the Multidimensional Fatigue Inventory (MFI; items 7, 11, 13, & 19; Smets, Garssen, Bonke, & De Haes, 1995). Participants rated four statements (e.g., "It takes a lot of effort to concentrate right now") using a 7-point-scale (1 = *strongly disagree* to 7 = *strongly agree*; α = .78-.89; for descriptive statistics see Table 1). Mental fatigue was assessed at four time points: before vs. after the first counting task, and before vs. after the second counting task.

Manipulation check. After each counting task, participants provided answers to the questions “How hard did you have to concentrate during the counting task?” and “How exhausting was it to work on the just finished counting task?” as compared to a math exam of 30 minutes. Answers were anchored on 7-point-scales (1 = *strongly disagree* to 7 = *strongly agree*; $\rho_{\text{1st section}}$[2] = .90; $\rho_{\text{2nd section}}$ = .91; for descriptive statistics see Table 1).

Demographic and control questions. Final questions inquired demographic details, subjective effects of the within-persons design, and potential strategies that participants used in the Stroop task.

Results

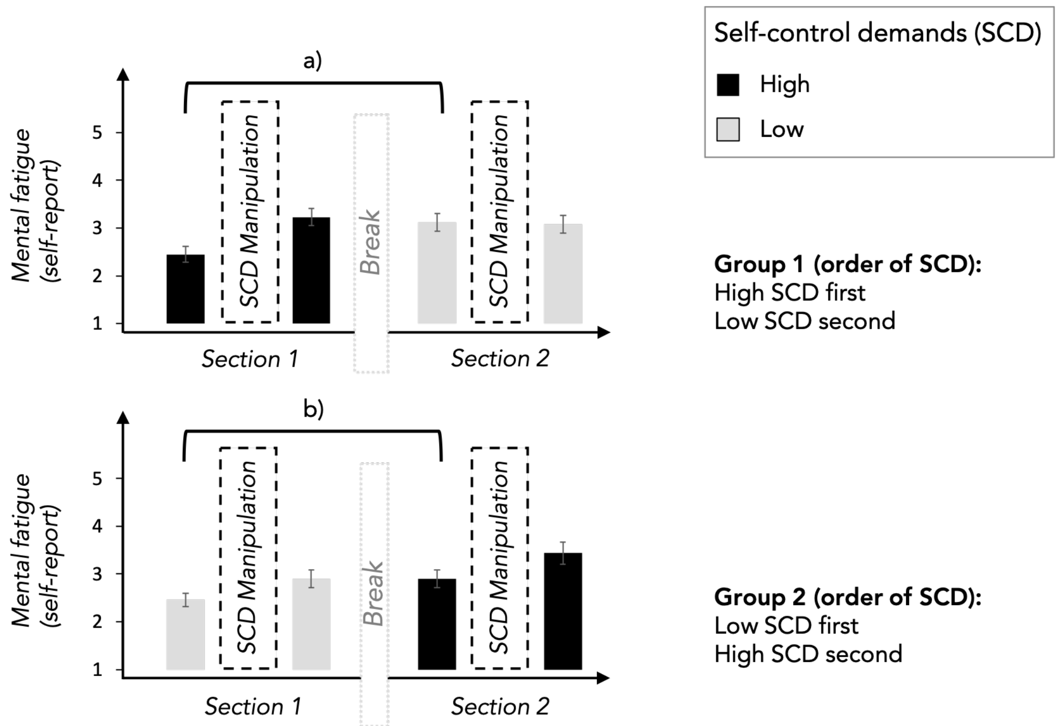
We report results as pre-registered if not mentioned otherwise. For the sake of readability, we report descriptive, frequentist inferential, and Bayesian statistics (JASP Team, 2019) relevant to the manipulation check, mental fatigue ratings, and the within-persons effect in Tables 1 and 2.

Preliminary note. Following the pre-registered criteria, three participants were excluded because their Stroop interferences based on error rates deviated more than 2.5 *SD*s from their condition mean.

Manipulation check. We analyzed manipulation check ratings in the context of a 2×2 mixed ANOVA with Self-control demands (*high* / *low*; within persons) and Task order (*high SCD first* / *low SCD first*; between persons). As expected, participants experienced the *highly self-control demanding* version of the counting task as more demanding than the *low self-control demanding* version. Neither the Task order main effect, nor the interaction between Self-control demands and Task order were significant (Table 1).

[2] Spearman-Brown reliability coefficient, recommended for two item scales (Eisinga, Grotenhuis, & Pelzer, 2013).

Mental fatigue. A 2×2×2 mixed ANOVA with Self-control demands (*high* / *low*) and Point in time (*before* / *after* the counting task) as within-persons factors, and Task order as between-persons factor (*high SCD first* / *low SCD first*) revealed a marginally significant three-way interaction. Both two-way interactions for Self-control demands × Task order, and for Self-control demands × Point in time were significant. Participants reported higher mental



fatigue after the recreational period than at the beginning of the study independent of their Task order condition, suggesting no full recovery ($t[88] = -4.41$; $p < .001$, $d = -0.47$, 95%CI [-0.69, -0.25]; Figure 1, Table 1).

Figure 1. Self-reported mental fatigue as a function of Self-control demands, Point in time and Task order in Study 1. Error bases indicate ± 1 SEM. ED = Ego Depletion. 7-point scale.

Table 1

Descriptive and inferential statistics relevant to manipulation check and mental fatigue in Study 1

1A) Counting task – Manipulation check.

Descriptive statistics

		Task order		
		High SCD first ($N = 48$)	Low SCD first ($N = 41$)	*Overall*
Self-control demands	High	6.06 (1.07)	6.22 (0.81)	*6.14 (0.96)*
	Low	3.13 (1.45)	3.16 (1.44)	*3.14 (1.43)*
	Overall	*4.60 (1.94)*	*4.69 (1.93)*	

Inferential statistics

	$F(1,85)$	p	η_p^2	90% CI	BF_{10}[a]
Self-control demands	337.39	< .001	.799	[.730, .834]	1.24×10^{37}
Task order	0.22	.637	.003	[0, .046]	
Self-control demands × Task order	0.15	.700	.002	[0, .041]	

1B) Mental fatigue by Task order.

Descriptive statistics

		High SCD first		**Low SCD first**	
		Point in time		Point in time	
		Before counting task	After counting task	Before counting task	After counting task
Self-control demands	High	2.45 (1.06)	3.23 (1.21)	2.90 (1.21)	3.44 (1.48)
	Low	3.12 (1.24)	3.08 (1.33)	2.46 (0.89)	2.90 (1.15)

Inferential statistics

	$F(1,85)$	p	η_p^2	90% CI	BF_{10}[a]
Self-control demands	1.19	.279	.014	[0, .078]	
Point in time	22.93	< .001	.212	[.093, .324]	
Task order	0.06	.814	.001	[0, .029]	
Self-control demands × Point in time	6.22	.015	.070	[.007, .163]	
Self-control demands × Task order	12.68	< .001	.130	[.037, .237]	
Point in time × Task order	0.48	.492	.006	[0, .057]	
Self-control × Task order × Point in time	3.93	.051	.044	[0, .130]	1.15[b]

Note. M (*SD*). Counting Task – Manipulation Check: Self-reported subjective difficulty, 2 items. 7-point scale with higher ratings indicating higher difficulty. Mental Fatigue by Task order: 4 items adapted from Multidimensional Fatigue Inventory. 7-point scale with higher ratings indicating higher mental fatigue.
[a]Analysis conducted in JASP (JASP Team, 2019) using default options; *r* scale fixed effect = 0.5, *r* scale random effects = 1, *r* scale covariates = 0.354. [b]BF for interaction term: Null model including everything except interaction.

Main analyses.

Within-persons ego depletion effect based on Stroop error rates. A 2×2 mixed ANOVA with Self-control demands (*high* / *low*) and Task order (*high SCD first* / *low SCD* *first*) revealed that after the difficult counting task (*high SCD*) participants did not show significantly impaired performance as compared to after the easy counting task (*low SCD*). Neither the Task order main effect, nor the interaction between Self-control demands and Task order were significant. A Bayesian ANOVA conducted in JASP (JASP Team, 2019) revealed anecdotal support for the null hypothesis[3]. Stroop interferences correlated significantly between measurement occasions, $r(87) = .31$, $p = .003$.

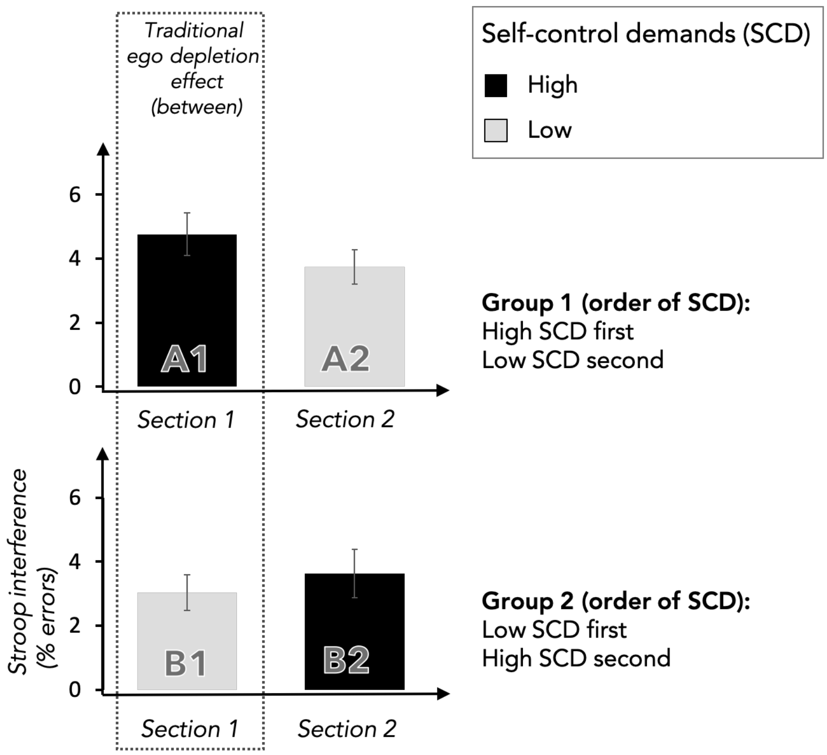


Figure 2. Stroop interference based on errors in Study 1. Error bars indicate ± 1 SEM. The order of self-control demands (high SCD first versus low SCD first) was the between participants factor. Self-control demands was the within participants factor (high versus low). Section 1 represents the traditional ego depletion effect between participants. The within-participants ego depletion effect compares the respective sections of high SCD [A1 / B2] versus low SCD [A2 / B1].

Within-persons ego depletion effect based on Stroop reaction times. The same 2×2 mixed ANOVA on reaction-times in the Stroop task revealed no main effect of Self-control

[3] By convention, a Bayes Factor (BF) provides anecdotal (1-3), moderate (3-10), strong (10-30), very strong (30-100) or extreme (> 100) evidence for a hypothesis (Schönbrodt & Wagenmakers, 2018).

demands: Participants did not show significantly impaired performance in the *high SCD* condition based on reaction times. There was a significant interaction, which can be interpreted as additional evidence for a learning effect: participants reacted faster in the second section of the study compared to the first section. Stroop interferences based on reaction times (r[87] = .61, $p < .001$) correlated significantly between measurement occasions. For descriptive and inferential statistics see Table 2.

Between-persons ego depletion effect based on Stroop error rates*[4]*. The traditional ego depletion effect between persons was significant, $t_{\text{one-tailed}}(85.25) = 2.26$; $p = .013$, d = 0.47, 95%CI [.11, +∞]. Participants in the *high SCD* condition performed significantly poorer (M = 4.96, SD = 4.74) than participants in the *low SCD* condition (M = 2.98, SD = 3.49). The Bayes factor suggested moderate evidence in favor of the ego depletion hypothesis (BF_{10} = 7.81, informed normal prior based on the assumed effect size of d = 0.35, M = 0.35, SD = 0.175, Dienes, 2014). To check for the robustness of this effect, we also used the default prior in JASP (Cauchy distribution, width = 0.707). This resulted in BF_{10} = 3.58. Note that informed priors are preferable to default priors because they can be based on theory and/or previous evidence (Dienes, 2014).

Between-persons ego depletion effect based on Stroop reaction times*[4]*. There was no significant ego depletion effect between participants on Stroop interference based on reaction times in the first section of the study ($M_{\text{highSCDfirst}}$ = 139.50, $SD_{\text{highSCDfirst}}$ = 92.62, $M_{\text{lowSCDfirst}}$ = 123.14, $SD_{\text{lowSCDfirst}}$ = 92.84; t[84.78] = .83, p = .205, d = 0.18, 95%CI [-.18, .53]). The estimated Bayes Factor (BF_{10} = 0.47; Cauchy prior width: 0.707) prompts the data to

[4] In contrast to Study 2, we did not preregister this as a directional hypothesis. This was an honest mistake. Following the plethora of ego-depletion studies and the lack of reverse depletion effects, we expected participants in the high SCD condition to perform worse, not better, than in the low SCD condition. Please note that descriptive statistics slightly differ compared to those for the within-participants effect reported in Table 2 (for Study 1) and Table 4 (for Study 2), because outlier exclusion for the between-participants effect is based on the first section of the study only.

moderately support the null hypothesis. This result is in line with previous research using a similar version of the Stroop task (Friese et al., 2013; Luethi et al., 2016).

Table 2

Descriptive and inferential statistics relevant to the within-persons effects in Study 1

1C) Stroop interference based on error rates.

Descriptive statistics

		Task order		
		High SCD first (N = 47)	Low SCD first (N = 40)	*Overall*
Self-control demands	High	4.76 (4.58)	3.63 (4.78)	*4.24 (4.68)*
	Low	3.74 (3.71)	3.03 (3.52)	*3.42 (3.62)*
	Overall	*4.25 (4.18)*	*3.33 (4.18)*	

Inferential statistics

	$F(1,85)$	p	η_p^2	90% CI	$p_{augmented}$	BF_{10}[a]
Self-control demands	2.36	.128	.027	[0, .087]	[.089, .147]	0.51
Task order	1.58	.211	.018	[0, .103]		
Self-control demands × Task order	0.15	.695	.002	[0, .041]		

1D) Stroop interference based on reaction times.

Descriptive statistics

		Task order		
		High SCD first (N = 47)	Low SCD first (N = 40)	*Overall*
Self-control demands	High	139.88 (93.59)	91.65 (65.45)	*117.71 (84.92)*
	Low	101.48 (89.91)	125.45 (92.83)	*112.50 (91.52)*
	Overall	*120.68 (93.62)*	*108.55 (81.60)*	

Inferential statistics

	$F(1,85)$	p	η_p^2	90% CI
Self-control demands	.08	.782	.001	[0, .033]
Task order	.53	.470	.006	[0, .059]
Self-control demands × Task order	18.94	< .001	.182	[.071, .293]

Note. *M* (*SD*). Stroop interference based on error rates: relative number of errors on congruent and incongruent trials (in %). Stroop interference based on reaction times: relative reaction difference in time on congruent and incongruent trials (in ms).
[a]Analysis conducted in JASP (JASP Team, 2019) using default options; *r* scale fixed effect = 0.5, *r* scale random effects = 1, *r* scale covariates = 0.354.

Discussion

Study 1 found no significant ego depletion effect *within* persons. This may be due to several reasons: First, statistical power to detect an assumed effect of $f = 0.175$ barely met the minimum of 80%, despite collecting more than the planned minimum sample size. This was due to the correlation between measurement occasions being lower than expected (r = .31). Second, the recreational period after the first section of the study failed to fully re-establish pre-study conditions: Independent of the task order, participants reported more mental fatigue after the recreational period than at the beginning of the study. Third, 14 participants (15%) reported using strategies circumventing the self-control demands of the Stroop task, effectively making the task easier. For instance, they reported not looking at the presented words but at the periphery of the screen to avoid reading and thus suppressing the common Stroop interference. Quite possibly, more participants used such strategies without reporting them.

Study 2 addressed these issues to provide a second and more comprehensive test of the possibility to investigate ego depletion within-persons. To ameliorate the concerns above, we (a) recruited a larger sample based on a more conservative power calculation, (b) modified the recreational period in order to make it more relaxing, and (c) created a modified version of the Stroop task to prevent the use of simplifying strategies that circumvent the exertion of self-control. In addition, we used a different measure of mental fatigue: As the MFI items primarily focused on concentration (e.g., "It takes a lot of effort to concentrate right now"), we opted for a measure stressing subjective mental fatigue over concentration.

Study 2

Method

Participants and design. We based our a-priori sample size calculation on the empirical correlation between measurement occasions in Study 1 (r[85] = .34)[5], the empirical effect size in Study 1 of f = 0.157, α = 0.05, and a power of 1-β = .90. This led to a minimum required sample size of N = 144. As this is a large, resource-intensive sample size, we pre-

[5] Based on an initial analysis erroneously excluding two more outliers.

registered using Bayes Factor Design Analysis (BFDA, Schönbrodt & Wagenmakers, 2018). BFDA allows researchers to not collect the complete sample, if an a-priori determined evidential threshold is met before. We followed other researchers (Schönbrodt, 2016) and used as evidential thresholds $BF_{10} = 6$ (for the alternative hypothesis) and $BF_{10} = 1/10$ for the null hypothesis. Once an initial sample of $N = 80$ was collected, we decided after each additional $N = 10$ about the continuation of data collection. As neither BF threshold was met before reaching 144 participants, we collected the full, pre-registered sample.

Participants were randomly assigned to a 2 (Self-control demands: *high* vs. *low*) × 2 (Task order: *high SCD first* vs. *low SCD first*) mixed design with the first factor varied within persons. Participants from Study 1 and psychology students were excluded from participation. One participant canceled the study because of dyscalculia and the data of seven additional participants were recorded incompletely due to technical problems, leading to a final sample of $N = 136$ ($M_{age} = 22.01$, $SD = 2.83$, 99 females).

Procedure.

Ego depletion manipulation. This manipulation was the same as in Study 1.

Recreational period. We modified the recreational period to make it more relaxing. Participants chose one of three relaxing pieces of music (i.e., Piano: *Comptine d'un autre été*, Yann Tiersen; Jazz: *Country*, Keith Jarrett; Lounge: *Northern Lights*, Lux). They listened to the music for approximately five minutes and received instructions to relax without further stimulus input for another three minutes. We expected this period to counteract depletion due to (a) the total amount of time that passed (Tyler & Burns, 2008), (b) relaxation through music and explicit instructions to relax (Tyler & Burns, 2008), and (c) autonomy in choice of music (Muraven, Gagné, & Rosman, 2008).

Measures.

Stroop task. The Stroop task was similar to the one used in Study 1 with the following exceptions: First, there were only three instead of four color words so that participants used three fingers of one hand instead of two fingers of both hands to answer. Second, we adapted the Stroop task to prevent participants from using strategies that circumvent the Stroop effect: Whenever the word “yellow” appeared, participant had to indicate the semantic meaning of the word instead of its font color. This rule forces participants to read the presented stimuli and prevents them from completing the task by looking at the periphery of the screen without reading the words. Third, a warning message was shown prompting for a quicker reaction, if a participant did not provide a response after 2000ms. Prompting participants to react quickly was expected to boost the effect on error rates. Fourth, each Stroop task consisted of 4 blocks comprising 45 trials each (56% congruent, 28% incongruent, 16% “yellow” exception rule). As for 30 participants only the first three blocks could be analyzed due to a programming error, we based our analyses on the first 3 blocks for all participants. Analyzing the data of all four blocks led to the same conclusions. Fifth, the practice trials were presented at the beginning of the experiment before the first counting task.

Questionnaires. The questionnaires were identical to Study 1 with one exception: We assessed mental fatigue with five items of the State Self-Control Capacity Scale (Ciarocco, Twenge, Muraven, & Tice, 2007).

Results

For descriptive, frequentist inferential, and Bayesian statistics on manipulation check, mental fatigue ratings, and the within-persons effect see Table 3 and 4.

Preliminary note. Following the pre-registered criterion, ten participants were excluded due to overall latency Stroop trials or overall number of congruent errors. Six participants were excluded because their Stroop interferences based on errors deviated more than 2.5 *SD*s from their condition mean. Four participants were excluded from the analyses based on reaction times following the pre-registered outlier criterion (> 2.5 *SD*s from individual mean reaction time; M_{rt} based on all correct trials; for analyses pertaining to reaction times).

Manipulation check. A 2×2 mixed ANOVA with Self-control demands (*high* / *low*; within persons) and Task order (*high SCD first* / *low SCD first*; between persons) revealed that, as expected, self-reported subjective difficulty of the counting task was higher in the *high SCD* as compared to the *low SCD* condition. Task order had an unexpected significant effect, with participants in the *high SCD first* condition generally experiencing both counting tasks as easier than participants in the *low SCD first* condition. There was no significant interaction between Self-control demands and Task order.

Mental fatigue. Similar to Study 1, we ran a 2×2×2 mixed ANOVA with Self-control demands (*high* / *low*) and Point in time (*before* / *after* the counting task) as within-persons factors and Task order as between-persons factor (*high SCD first* / *low SCD first*). The three-way interaction was non-significant. The two-way interactions for Self-control demands × Point in time and Point in time × Task order were significant (Table 3). Concerning mental fatigue *before* each counting task, participants in the *high SCD first* condition reported higher mental fatigue *after* the recreational period than at the beginning, again suggesting no full recovery ($t_{highSCDfirst}[57] = -2.20$, $p = .032$, $d = -0.29$, 95%CI [-0.55, -.03]; $t_{lowSCDfirst}[63] = -1.40$, $p = .167$, $d = -0.17$, 95%CI [-0.42, 0.07]; Figure 3).

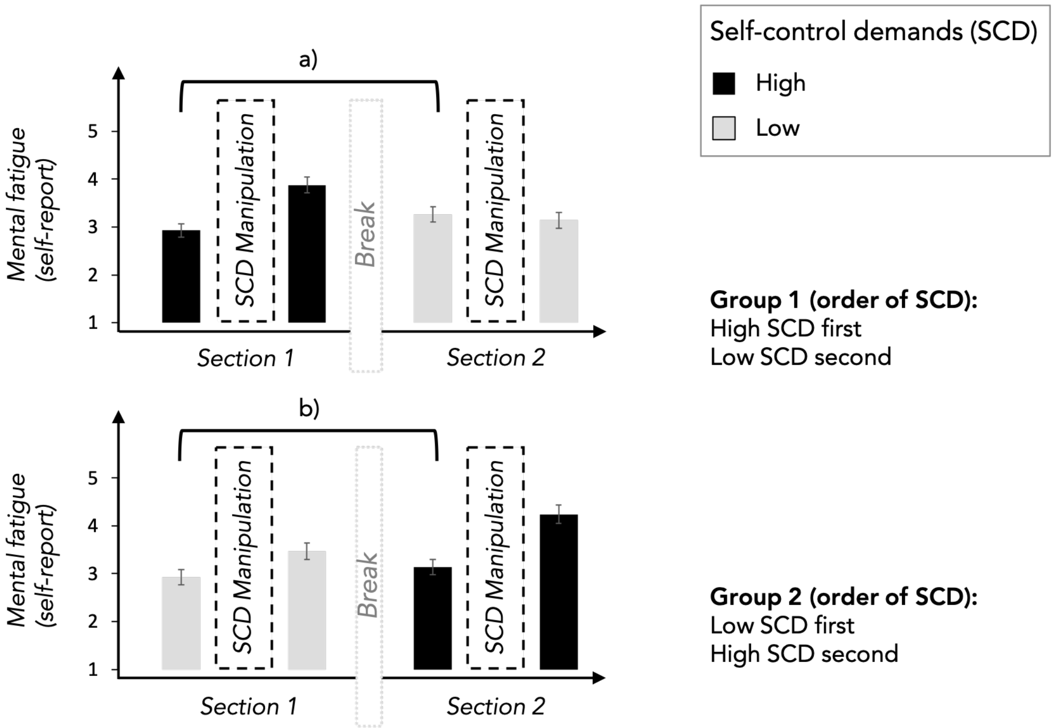


Figure 3. Self-reported mental fatigue as a function of Self-control demands, Point in time and Task order in Study 2. Error bases indicate ± 1 SEM. ED = Ego Depletion. 7-point scale.

Table 3

Descriptive and inferential statistics relevant to manipulation check and mental fatigue in Study 2

2A) Counting task – Manipulation check.

Descriptive statistics

		Task order		
		High SCD first (N = 58)	Low SCD first (N = 64)	*Overall*
Self-control demands	High	5.89 (1.23)	6.24 (1.07)	*6.07 (1.16)*
	Low	2.83 (1.49)	3.74 (1.66)	*3.31 (1.64)*
	Overall	*4.36 (2.05)*	*4.99 (1.88)*	

Inferential statistics

	$F(1,120)$	p	η_p^2	90% CI	BF_{10}[a]
Self-control demands	330.52	< .001	.078	[.024, .210]	7.75×10^{38}
Task order	10.15	< .001	.734	[.726, .831]	
Self-control demands × Task order	3.36	.069	.027	[0, .120]	

2B) Mental fatigue by Task order.

Descriptive statistics

		High SCD first		**Low SCD first**	
		Point in time		Point in time	
		Before counting task	After counting task	Before counting task	After counting task
Self-control demands	High	2.93 (1.06)	3.88 (1.33)	3.14 (1.28)	4.24 (1.49)
	Low	3.26 (1.21)	3.14 (1.24)	2.93 (1.30)	3.47 (1.35)

Inferential statistics

	$F(1,120)$	p	η_p^2	90% CI	BF_{10}[a]
Self-control demands	18.39	< .001	.133	[.068, .288]	
Point in time	70.37	< .001	.370	[.317, .545]	
Task order	0.55	.461	.005	[0, .060]	
Self-control demands × Point in time	32.70	< .001	.214	[.147, .387]	
Self-control demands × Task order	3.04	.084	.025	[0, .115]	
Point in time × Task order	7.43	.007	.058	[.012, .178]	
Self-control × Task order × Point in time	3.05	.083	.025	[0, .115]	0.50[b]

Note. *M* (*SD*). Counting Task – Manipulation Check: Self-reported subjective difficulty, 2 items. 7-point scale with higher ratings indicating higher difficulty. Mental Fatigue by Task order: 5 items out of the State Self-Control Capacity Scale. 7-point scale with higher ratings indicating higher mental fatigue.
[a]Analysis conducted in JASP (JASP Team, 2019) using default options; r scale fixed effect = 0.5, r scale random effects = 1, r scale covariates = 0.354. [b]BF for interaction term: Null model including everything except interaction.

Main analyses.

Within-persons ego depletion effect based on Stroop error rates. A 2×2 mixed ANOVA with Self-control demands (*high* / *low*; within persons) and Task order (*high SCD first* / *low SCD first*; between persons) revealed that Stroop interferences were generally stronger in the *high SCD* condition than in the *low SCD* condition (see Figure 4). Importantly, this within-persons depletion main effect was qualified by a strong interaction between Self-control demands and Task order. In the *high SCD first* condition, Stroop performance was poorer after high Self-control demands (first section) as compared to after low Self-control demands in the counting task (second section, $t_{\text{two-tailed}}[58] = 5.28$; $p < .001$, $d = 0.69$, 95%CI [0.40, 0.97]), but in the *low SCD first* condition, this effect was reversed ($t_{\text{two-tailed}}[63] = -4.17$; $p < .001$, $d = -0.52$, 95%CI [-0.78, -0.26]). In essence, there was a strong learning effect: Performance was always better for the second than for first Stroop task ($t[122] = 6.53$; $p < .001$, $d = 0.59$, 95%CI [0.40, 0.78]). The correlation of Stroop interferences between measurement occasions was moderate ($r_{\text{overall}}[121] = .37$, $p < .001$); it differed somehow between the two independent subsamples ($r_{\text{highSCDfirst}}[57] = .30$, $p = .024$; $r_{\text{lowSCDfirst}}[62] = .55$, $p < .001$).

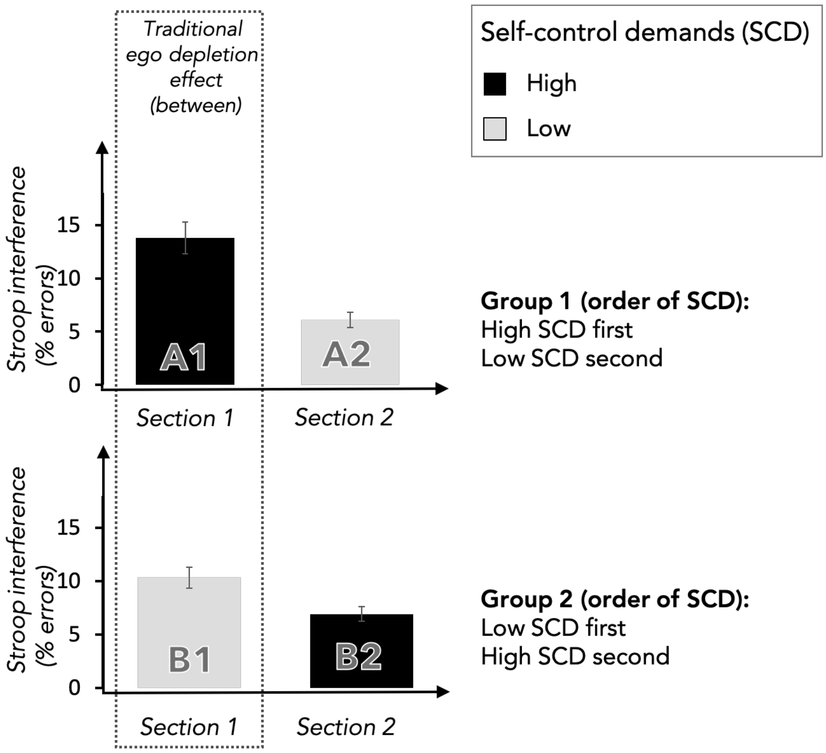


Figure 4. Stroop interference based on errors in Study 2. Error bars indicate ± 1 SEM. The order of self-control demands (high SCD first versus low SCD first) was the between participants factor. Self-control demands was the within participants factor (high versus low). Section 1 represents the traditional ego depletion effect between participants. The within-participants ego depletion effect compares the respective sections of high SCD [A1 / B2] versus low SCD [A2 / B1].

Within-persons ego depletion effect based on Stroop reaction times. Similar to Study 1, a 2×2 repeated-measures ANOVA with Self-control demands (*high* / *low*; within participants) and Task order (*high SCD first* / *low SCD first*; between participants) on Stroop reaction times revealed that participants did not differ significantly as a function of Self-control demands or Task order. However, there was a significant interaction underlining the learning effect found for Stroop interference on errors: Participants' Stroop interference on reaction times decreased from the first to the second section (Table 4). Stroop interferences based on reaction times correlated significantly between measurement occasions ($r_{\text{overall}}[123] = .60$, $p < .001$; $r_{\text{highSCDfirst}}[58] = .61$, $p < .001$; $r_{\text{lowSCDfirst}}[63] = .61$, $p < .001$).

***Between-person ego depletion effect based on Stroop error rates*[6].** Looking at the first experimental section only, participants in the *high SCD first* condition showed poorer Stroop performance compared to participants in the *no high SCD first* condition (see bars 1 and 4 in Figure 4). This provides evidence for the traditional ego depletion effect *between* persons ($M_{\text{highSCDfirst}} = 14.03$, $SD_{\text{highSCDfirst}} = 11.45$; $M_{\text{lowSCDfirst}} = 10.59$, $SD_{\text{lowSCDfirst}} = 7.72$; $t_{\text{one-tailed}}[103.83] = 1.98$; $p = .025$, $d = 0.36$, 95%CI = [.06, + ∞]). The estimated Bayes Factor ($BF_{10} = 5.39$; informed normal prior: $M = 0.35$, $SD = 0.175$) suggested that the data "moderately" support the ego depletion hypothesis. Checking for the robustness of this effect with the default prior in JASP, we obtained $BF_{10} = 2.28$ (Cauchy prior width = 0.707).

***Between-person ego depletion effect based on Stroop reaction times*[6].** Again, there were no differences in Stroop interference based on reaction times depending on Self-control demands ($M_{\text{highSCDfirst}} = 200.78$, $SD_{\text{highSCDfirst}} = 70.98$, $M_{\text{lowSCDfirst}} = 205.57$, $SD_{\text{lowSCDfirst}} = 77.43$; $t(126.94) = -0.37$; $p = .715$, $d = -.06$, 95%CI [-0.36, 0.23]). The estimated Bayes Factor ($BF_{10} = 0.200$; Cauchy prior width: 0.707) prompts the data to moderately support the *null* hypothesis. This is in line with previous research (Friese et al., 2013; Luethi et al., 2016) and Study 1.

[6] Please note that descriptive statistics slightly differ compared to those for the within-participants effect reported in Table 4 (for Study 2), because outlier exclusion for the between-participants effect is based on the first section of the study only.

Table 4

Descriptive and inferential statistics relevant to the within-persons effect in Study 2

2C) Stroop interference based on errors.

Descriptive statistics

		Task order		
		High SCD first	Low SCD first	*Overall*
Self-control demands	High	13.76 (11.48)	10.32 (7.65)	*10.20 (9.51)*
	Low	6.10 (5.36)	6.91 (5.55)	*8.29 (6.95)*
	Overall	*9.93 (9.71)*	*8.62 (6.87)*	

Inferential statistics

	$F(1,121)$	p	η_p^2	90% CI	BF_{10}[a]
Self-control demands	6.84	.010	.054	[.010, .171]	1.01
Task order	1.28	.259	.011	[0, .080]	
Self-control demands × Task order	46.05	< .001	.276	[.214, .455]	3.12×10^{7} [b]

2D) Stroop interference based on reaction time.

Descriptive statistics

		Task order		
		High SCD first	Low SCD first	*Overall*
Self-control demands	High	199.74 (71.11)	187.67 (55.21)	*193.47 (63.37)*
	Low	170.77 (57.25)	203.44 (75.85)	*187.76 (69.27)*
	Overall	*185.25 (65.91)*	*195.56 (66.55)*	

Inferential statistics

	$F(1,123)$	p	η_p^2	90% CI
Self-control demands	1.53	.219	.013	[0, .086]
Task order	0.97	.326	.008	[0, .072]
Self-control demands × Task order	17.56	< .001	.125	[.063, .281]

Note. M (*SD*). Stroop interference based on errors: relative number of errors on congruent and incongruent trials (in %). Stroop interference based on reaction times: relative reaction difference in time on congruent and incongruent trials (in ms).
[a]Analysis conducted in JASP (JASP Team, 2019) using default options; *r* scale fixed effect = 0.5, *r* scale random effects = 1, *r* scale covariates = 0.354. [b]BF for interaction term: Null model including everything except interaction.

Discussion

The first goal of Study 2 was higher statistical power than in Study 1. This goal was achieved. Power was 93% assuming a true effect of $f = 0.175$, $\alpha = 0.05$ and the empirical correlation based on the complete sample of $r = .37$ between measurements.

The second goal was to make the recreational period more relaxing. This was successful for participants in the *low SCD first* condition; but for participants in the *high SCD* *first* condition, the recreational period did not lead to full recovery. Third, we aimed at preventing participants from using strategies to circumvent the Stroop's self-control demands. We therefore developed a modified version of the task, forcing participants to read the words instead of circumventing the Stroop effect by looking at the periphery of the screen. Overall stronger Stroop interferences as compared to Study 1 suggest that this goal was achieved. In addition, this underlines the appropriateness of the Stroop task as an indicator of self-control that forces participants to control their prepotent responses.

Study 2 showed a within-persons ego depletion effect, but only when participants started with the highly self-control demanding section of the study. In the other Task order condition starting with the low-demanding section of the study, a reverse-depletion effect occurred. Together, these data strongly suggest a pronounced learning effect: Performing the Stroop task for the second time always led to markedly weaker interference scores than performing the task for the first time. Finally, Study 2 again found a traditional between-persons ego depletion effect for the first experimental section.

General Discussion

The aim of the present research was to introduce and examine the feasibility of a within-persons design in ego depletion research that allows for higher statistical power than the traditional between-persons design. A subordinate goal was to examine the traditional between-persons effect in a pre-registered format. Regarding the former goal, the data indicates that ego depletion cannot be examined easily with the dual task paradigm twice within the same experimental session—at least not with the independent and dependent variables we employed here. Considering our subordinate goal, both studies delivered pre-registered evidence for the heavily contested traditional ego depletion effect between persons (no file drawer, public materials and data).

Despite the significant within-persons effect, Study 2 does not provide convincing evidence for the feasibility to manipulate ego depletion within persons in a single experimental session. In both Task order conditions, there was a strong learning effect such that Stroop performance in the second section was better than in the first section—irrespective of whether the second Stroop task was performed after the counting task high or low in self-control demands. Thus, future research should investigate other dependent variables that are potentially less likely to show similarly strong learning effects or employ dependent variables that lend themselves to the development of non-identical but parallel versions for each experimental section. Additionally, effects concerning mental fatigue should be considered: Despite efforts to make the recreational period as relaxing as possible, participants generally reported more pronounced mental fatigue after the resting period than at the beginning of the study.

The reader may wonder if these two manifestations of the ego depletion effect are particularly noteworthy after hundreds of studies have supported the idea of ego depletion. Indeed, for anyone convinced that ego depletion is real, two further studies supporting the effect may not be exciting news. However, after a meta-analysis finding "very little evidence that the depletion effect is a real phenomenon" (Carter et al., 2015, p. 796) and a large-scale RRR finding a null effect (Hagger & Chatzisarantis, 2016), a new consensus seems to have emerged suggesting that ego depletion is *not* real. This view asserts that supportive evidence is the sole result of publication bias (Carter et al., 2015; Carter & McCullough, 2014) and questionable research practices. Indeed, the presence of publication bias and questionable research practices in the ego depletion literature is a reality that can hardly be denied, similar to many other fields (Friese et al., 2019). In our view, the only way to convince skeptical researchers (including ourselves) that there is something to say about ego depletion, is to provide convincing empirical evidence: Pre-registered studies with open materials, open data, and no file drawer. From this perspective, we believe that the present two studies are noteworthy. So far, there is little pre-registered research to reveal evidence in favor of the ego depletion effect (Dang, Liu, Liu, & Mao, 2017; Garrison, Finley, & Schmeichel, 2019). Clearly, more pre-registered work providing further replications and using other combinations of independent and dependent variables is needed to gain a better understanding of the robustness and boundary conditions of the ego depletion effect. We examined only one combination of (in)dependent variables and it is needless to say that many more need to be examined in an open and transparent manner to change vigorously held beliefs of any kind whatsoever (Carter & McCullough, 2018)—not to mention a thorough theoretical occupation with possible underlying mechanisms.

Conclusion

The existence of the ego depletion effect has recently been widely questioned, suggesting that – if at all – large sample sizes would be needed to find evidence for this effect using a traditional between-participants dual task paradigm. We aimed at finding a parsimonious way of testing the ego depletion effect in a more efficient within-persons design. We were not successful in this endeavor as our within-persons manipulation did not work, presumably due to strong learning effects. Somewhat ironically, however, the data of both studies provide pre-registered, open, and transparent evidence in favor of an ego depletion effect *between* persons.

References

Baumeister, R. F., & Vohs, K. D. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*, 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x

Baumeister, R. F., & Vohs, K. D. (2016). Strength model of self-regulation as limited resource: Assessment, controversies, update. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 54, pp. 67–127). San Diego, CA: Academic Press. https://doi.org/10.1016/bs.aesp.2016.04.001

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

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. http://dx.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., & McCullough, M. E. (2018). A simple, principled approach to combining evidence from meta-analysis and high-quality replications. *Advances in Methods and Practices in Psychological Science*, *1*, 174–185. https://doi.org/10.1177/2515245918756858

Ciarocco, N. J., Twenge, J., Muraven, M., & Tice, D. M. (2007). *State Self-Control Capacity Scale*. Monmouth, NJ: Monmouth University. Unpublished manuscript.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: Lawrence Erlbaum Associates. Inc. Publishers.

Converse, P. D., & Deshon, R. P. (2009). A tale of two tasks: Reversing the self-regulatory resource depletion effect. *The Journal of Applied Psychology*, *94*, 1318–1324. https://doi.org/10.1037/a0014604

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

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology*, *5*, 781. https://doi.org/10.3389/fpsyg.2014.00781

Eisinga, R., Grotenhuis, M. te, & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637–642. https://doi.org/10.1007/s00038-012-0416-3

Faul, F., Erdfelder, E., Lang, 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. https://doi.org/10.3758/BF03193146

Friese, M., Binder, J., Luechinger, R., Boesiger, P., & Rasch, B. (2013). Suppressing emotions impairs subsequent Stroop performance and reduces prefrontal brain activation. *PLoS ONE*, *8*, e60385. https://doi.org/10.1371/journal.pone.0060385

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

Garrison, K. E., Finley, A. J., & Schmeichel, B. J. (2019). Ego depletion reduces attention control: Evidence from two high-powered preregistered experiments. *Personality and Social Psychology Bulletin*, *45*, 728–739. https://doi.org/10.1177/0146167218796473

Graham, J. D., & Bray, S. R. (2015). Self-control strength depletion reduces self-efficacy and impairs exercise performance. *Journal of Sport and Exercise Psychology*, *37*, 477–488. https://doi.org/10.1123/jsep.2015-0064

Hagger, M. S., & Chatzisarantis, N. L. D. (2014). It is premature to regard the ego-depletion effect as "Too Incredible." *Frontiers in Psychology*, *5*, 298. https://doi.org/10.3389/fpsyg.2014.00298

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/

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/10.1177/1745691616652873

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

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes Factors. *The Journal of Problem Solving*, *7*, 2–9. https://doi.org/10.7771/1932-6246.1167

JASP Team (2019). JASP (Version 0.10.2)[Computer software].

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: Examining the influence of emotion regulation. *Journal of Experimental Psychology. General*, *137*, 691–705. https://doi.org/10.1037/a0013834

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The "Trier Social Stress Test" - A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*, 76–81. https://doi.org/10.1159/000119004

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

Muraven, M., Collins, R. L., & Nienhaus, 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., Gagné, M., & Rosman, H. (2008). Helpful self-control: Autonomy support, vitality, and depletion. *Journal of Experimental Social Psychology*, *44*, 573–585. https://doi.org/10.1016/j.jesp.2007.10.008

Sagarin, B. J., Ambler, J. K., & Lee, E. M. (2014). An ethical approach to peeking at data. *Perspectives on Psychological Science*, *9*, 293–304. https://doi.org/10.1177/1745691614528214

Sailer, S. (2011). *Wild Faces of the Andes*.

Schönbrodt, F. D. (2016). Pre-registered replication of Schiestel & Schönbrodt: "Does the experience of presence restore self-control?" Retrieved March 25, 2016, from osf.io/54s7m

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 1–15. https://doi.org/10.3758/s13423-017-1230-y

Smets, E. M. A., Garssen, B., Bonke, B., & De Haes, J. C. J. M. (1995). The multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research*, *39*, 315–325. https://doi.org/10.1016/0022-3999(94)00125-O

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

Tice, D. M., Baumeister, R. F., Shmueli, D., & Muraven, M. (2007). Restoring the self: Positive affect helps improve self-regulation following ego depletion. *Journal of Experimental Social Psychology*, *43*, 379–384. https://doi.org/10.1016/j.jesp.2006.05.007

Tyler, J. M., & Burns, K. C. (2008). After depletion: The replenishment of the self's regulatory resources. *Self and Identity*, *7*, 305–321. https://doi.org/10.1080/15298860701799997

Supplement

Questionnaires

In both studies, we assessed the following questionnaires in addition to those already mentioned in the manuscript.

Subjective effort and motivation during the Stroop task. Two questions assessed subjective effort during each Stroop task (e.g., "How difficult was it for you to process the color task in line with the instructions?", "How strenuous did it feel for you to work on the color task according to the instructions?", 7-point scales from 1 = *not at all* to 7 = *very much*). Three additional items assessed motivation during the Stroop task (e.g., "How motivated were you to work on the color task in line with the instruction?", "How much did you try to work on the color task in line with the instructions?", "How much did you feel like processing the color task in line with the instructions?", 7-point scales from 1 = *not at all* to 7 = *very much*).

Lay theories of willpower. Six items assessed participants' lay theories about willpower (Job, Dweck, & Walton, 2010). The items where embedded among other questionnaire items assessing implicit theories concerning intelligence, emotions, and temptations in order to mask the purpose of the questions. Participants indicated their approval to statements like "After a strenuous mental activity, I feel energized for further challenging activities" or "My mental stamina fuels itself. Even after strenuous mental exertion, I can continue doing more of it" on 6-point scales (1 = *strongly agree*; 6 = *strongly disagree*).

Participants' impression of the within-design. Participants rated their agreement on three statements concerning their motivation during the second experimental section ("During the second experimental section, I did no longer feel like doing similar tasks again.", "I thought of the second experimental section as a chance to improve my performance.", "In the second experimental section I was motivated to show that I can master the tasks well.", 6-point scales from 1 = *not at all* to 6 = *very much*).

On three additional statements, participants rated their agreement concerning their exhaustion during the second experimental section ("After the recreational period, I was more exhausted than before.", "After the recreational period, I felt as mentally refreshed as at the beginning of the study", "The recreational period helped me to recover from the first experimental section", 6-point scales from 1 = *not at all* to 6 = *very much*).

Demographic questions. We assessed participants' age, sex, profession, and German language skills.

Closing questions.

Data quality. We assessed, how participants judged the quality of their data ("Sometimes, participants get distracted during a study, for example because a phone rings, or do not take the study seriously. If this applies to you, we ask you to indicate this. How do you judge the quality of your data? a) My data is fine. b) I am not sure, whether my data is fine. c) The quality of my data is negatively affected for sure."). None of the participants answered 'c'. Excluding participants who answered 'b' (Study 1: $N = 4$; Study 2: $N = 7$) in addition to our pre-registered outlier criterion did not change the conclusions concerning the ego depletion effect *within* or *between* persons based on errors.

Cheering and relaxing effects of the recreational period. We assessed participants' impression of the recreational period (Study 1: "How did you perceive the interview with Roger Federer", 1 = *rather not funny* to 7 = *very funny*, "How did you perceive the landscape video?", 1 = *rather tensioning* to 7 = *very relaxing*; Study 2: "How did you perceive the piece of music?", 1

= *rather tensioning* to 7 = *very relaxing*, “How did you perceive the time to relax and settle back?”, 1 = *rather tensioning* to 7 = *very relaxing*).

Study 2

Pre-registered exploratory analyses.

Ego depletion effect within persons for outliers based on Median Absolute Deviations (Leys et al., 2013). During the planning phase for Study 2, we learned about an alternative method to identify outliers, Median Absolute Deviations (MAD; Leys et al., 2013). The MAD is based on deviation from the median instead of deviation from the mean. It has the advantage that it is not as much influenced by the outliers as is the deviation from the mean. Therefore, we applied this procedure in a separate exploratory analysis, as pre-registered on OSF.

Stroop interference based on errors. The results for outliers based on MAD are similar to those for outliers based on our standard outlier criterion (Stroop interference > 2.5 *SD* from group mean). Overall, the results show a strong learning effect: Performance was always better for the second than for the first Stroop task, as displayed by the strong interaction between Self-control demands and Task order. Stroop interference correlated significantly between measurement occasions ($r_{\text{overall}}[126] = .40$, $p < .001$; $r_{\text{EDfirst}}[58] = .40$, $p = .001$; $r_{\text{noEDfirst}}[66] = .45$, $p < .001$). For descriptive and inferential statistics see Table 1.

Table 1

Stroop interference based on errors

Descriptive statistics				
		Task order		
		High SCD first	Low SCD first	*Overall*
Self-control demands	High	14.80 (11.73)	6.97 (5.30)	*10.64 (9.70)*
	Low	6.19 (5.19)	11.10 (8.26)	*8.80 (7.39)*
	Overall	*10.49 (10.01)*	*9.04 (7.22)*	

Inferential statistics			
	F(1,126)	*p*	η_p^2
Self-control demands	7.62	.007	.057
Task order	1.57	.213	.012
Self-control demands × Task order	61.40	< .001	.328

Note. *M* (*SD*). Stroop interference: relative number of errors on congruent and incongruent trials (in %). Outliers based on Median Absolute Deviation (Leys et al., 2013).

Stroop interference based on reaction time. This MAD-based analysis revealed the same results as the one with outliers based on our standard criterion. There was a significant interaction with participants in the both conditions showing faster reaction times in the second section. Stroop interference correlated significantly between measurement occasions ($r_{overall}$[132] = .638, $p < .001$; $r_{EDfirst}$[63] = .57, $p < .001$; $r_{noEDfirst}$[71] = .69, $p < .001$). For descriptive and inferential statistics see Table 2.

Table 2

Stroop interference based on reaction time

Descriptive statistics

		Task order		
		High SCD first	Low SCD first	*Overall*
Self-control demands	High	194.95 (80.71)	182.62 (63.60)	*188.41 (72.14)*
	Low	164.39 (55.58)	201.33 (88.90)	*183.97 (77.09)*
	Overall	*176.67 (70.70)*	*191.97 (77.59)*	

Inferential statistics

	$F(1,132)$	p	η_p^2
Self-control demands	1.09	.299	.008
Task order	1.16	.283	.009
Self-control demands × Task order	18.81	< .001	.125

Note. M (*SD*). Stroop interference: relative reaction time on congruent and incongruent trials (in ms). Outliers based on Median Absolute Deviation.

Standard ego depletion effect between persons as a function of task order for outliers based on MAD (Leys et al., 2013).

Stroop interference based on errors. Similar to the results based on the standard outlier criterion, participants in the *high SCD first condition* showed poorer Stroop performance compared to participants in the *low SCD first condition* in the first section of the study, providing evidence for the standard ego depletion effect between persons (M_{EDfirst} = 14.83, SD_{EDfirst} = 11.81; $M_{\text{noEDfirst}}$ = 11.49, $SD_{\text{noEDfirst}}$ = 8.45; $t_{\text{one-tailed}}[109.08]$ = 1.85; p = .033, d = 0.33, 95%CI [.04, +∞]).

Stroop interference based on reaction time. Similar to the results based on the standard outlier criterion, the standard ego depletion effect between participants was not significant (M_{EDfirst} = 198.19, SD_{EDfirst} = 84.15, $M_{\text{noEDnfirst}}$ = 201.33, $SD_{\text{noEDfirst}}$ = 88.90; t (132.67) = -0.21; p = .833, d = -.04, 95%CI [-0.32, 0.25]).

References

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/10.1177/0956797610384745

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766. https://doi.org/10.1016/j.jesp.2013.03.013

PART II

A Pre-Registered Test of Competing Theories to Explain Ego Depletion Effects Using Psychophysiological Indicators of Mental Effort

Karolin Gieseler
Saarland University

David D. Loschelder
Leuphana University of Lueneburg

Veronika Job
Technical University Dresden

Malte Friese
Saarland University

Manuscript submitted to *Motivation Science*.
(Currently under revision)

Abstract

A prominent, hotly debated idea—the 'ego depletion' phenomenon—suggests that engaging in effortful, demanding tasks leads to poorer subsequent self-control performance. Several different theories seek to explain the emergence of ego depletion effects. The two most prominent ones are the strength model of self-control (Baumeister & Vohs, 2016) and the process model of self-control (Inzlicht & Schmeichel, 2012). Predictions of these models are predominantly identical on the behavioral level. The models' predictions differ, however, on the level of invested mental effort. The present pre-registered study (N = 179) contrasted these competing predictions combining an established moderator of ego depletion effects (i.e., self-affirmation) and psychophysiological indicators of mental effort (i.e., systolic blood pressure and pre-ejection period). Our data provide moderate evidence for ego-depletion—decrements in self-control performance after a *high-* versus *low*-demanding task in the *non-affirmed* conditions. Self-affirmation had an unexpected effect: Contrary to previous research, *self-affirmed* participants performed similarly poorly as participants in the *high demands+non-affirmed* condition. Although this finding limited the ability to contrast competing model predictions, it points to hitherto unknown effects of self-affirmation on self-control performance. Systolic blood pressure emerged as a valid indictor of invested mental effort; the data show no sign of disengagement after a high demanding task predicted by the process (but not the strength) model. We explore systolic blood pressure progression across the sequential task paradigm, suggest a verifiable account for the effects of self-affirmation on self-control performance, and discuss theoretical implications for the two competing models.

Keywords: self-control, ego depletion, mental effort, self-affirmation, construal level, pre-ejection period, psychophysiology, blood pressure

A pre-registered test of competing theories to explain ego depletion effects using psychophysiological indicators of mental effort

The literature on ego depletion is one of the most prolific and hotly debated literatures in psychology. Two prominent theoretical models—the strength model (Baumeister & Vohs, 2016) and the process model (Inzlicht & Schmeichel, 2012)—make predominantly identical predictions for self-control behavior; yet, these are based on radically different assumptions about *why* ego depletion effects occur. This makes it difficult to obtain empirical evidence that clearly speaks in favor of one over the other model. The present study goes beyond the observation of self-control performance and creates an experimental context that allows to contrast opposing predictions of the strength model versus the process model. To this end, we use the traditional ego depletion sequential task paradigm in combination with a) the moderator self-affirmation and b) psychophysiological measures of mental effort (i.e., systolic blood pressure and pre-ejection period) as a further layer of analysis in addition to behavioral self-control.

Ego Depletion – Support and Skepticism

Self-control is commonly conceptualized as the “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). A prominent hypothesis suggests that the initial exertion of self-control leads to the subsequent impairment of further self-control performance—the ‘ego depletion effect’ (Baumeister, Bratlavsky, Muraven, & Tice, 1998). For decades, the ego depletion idea has been examined in a sequential task paradigm, with the second task measuring self-control performance as a function of whether the preceding task was self-control demanding or not. Hundreds of studies provided support for the ego depletion idea and a first meta-analysis suggested that the ego depletion effect was a robust phenomenon

(Hagger, Wood, Stiff, & Chatzisarantis, 2010). However, a reanalysis of this meta-analysis and a more recent meta-analysis suggest the presence of severe publication bias in the ego depletion literature (Carter, Kofler, Forster, & McCullough, 2015; Carter & McCullough, 2014). A registered replication report found a null effect of one specific ego depletion study in a high-powered test (Hagger et al., 2016), and further recent studies also revealed null effects (e.g., Etherton et al., 2018; Lurquin et al., 2016; Osgood, 2017). All these and further criticisms gave rise to substantial doubts and critique, questioning the trustworthiness of extant ego depletion research and even putting the mere existence of the ego depletion phenomenon into question. Thus, at this stage, evidence in favor of ego depletion that meets current standards of rigor and openness is rare (but does exist, see Dang, Liu, Liu, & Mao, 2017; Garrison, Finley, & Schmeichel, 2019). With respect to the theoretical foundation of ego depletion, it appears striking that—even after hundreds of seemingly supportive studies—conclusive evidence for either the strength or the process model is lacking (for a review, see Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2019).

Theoretical Approaches

Multiple theoretical models converge in their behavioral predictions of ego depletion effects (functional level of analysis, De Houwer, 2011). The strength model (Baumeister & Vohs, 2016) and the process model (Inzlicht & Schmeichel, 2012) are the two most prominent of these models.[1] These models drastically differ on the cognitive level of analysis that strive to explain

---

[1] Alternative models are the opportunity cost model (Kurzban, Duckworth, Kable, & Myers, 2013), the labor/leisure tradeoff concept (Kool & Botvinick, 2014), or the central governor model (Evans, Boggero, & Segerstrom, 2015).

why ego depletion effects occur (De Houwer, 2011). We seek to disentangle the behavioral and psychophysiological predictions of these two models.

The strength model of self-control (Baumeister & Vohs, 2016; Baumeister, Vohs, & Tice, 2007) makes two basic assumptions: First, self-control is a domain general phenomenon and second, it relies on a limited (but otherwise not specified) resource that is reduced whenever a person engages in self-control. This means that exerting self-control in one domain reduces the availability of the resource and leads to the impairment of subsequent self-control performance in other domains as well (ego depletion idea). According to the strength model, situational factors may counteract mild depletion effects by mobilizing additional effort to recruit resources despite the general tendency conserve the precious and finite resource.

The process model (Inzlicht & Schmeichel, 2012; Inzlicht, Schmeichel, & Macrae, 2014) takes a different approach to explain behavioral ego depletion effects: It does not assume a limited resource. Instead, the exertion of self-control leads to a loss of motivation to further exert self-control—that is, after a demanding task people feel like doing something easier and more pleasurable. This is often described as a shift from “have-to goals” to “want-to goals”. In addition, attention and emotions change in line with the motivational shift. Thus, after exerting self-control people are still *able* but no longer motivated to exert further self-control. Hence, performance breaks down after a demanding task. The process model posits that it is possible to bring performance back to previous levels by reallocating motivation to self-control-demanding tasks.

In sum, since both models were developed to explain ego depletion, they both make the same prediction for the basic effect on the functional level of analysis in terms of observable behavior (De Houwer, 2011): The initial exertion of self-control leads to a subsequent decline in self-control performance. On this basis, it is difficult, if not impossible, to test the models' diverging predictions concerning the underlying mechanisms on the cognitive level of analysis (i.e., depleted resource versus shifts in motivation and attention). In the present study, we therefore extended the traditional sequential task paradigm to empirically contrast the two models: We included self-affirmation as a well-established moderator of ego depletion, *while* at the same time assessing psychophysiological indicators of mental effort that allow for an additional layer of analysis beyond behavioral self-control performance. In combination, these extensions allow us to differentiate between the models' theoretical assumptions, as we elucidate in the next two sections.

Self-Affirmation

One fundamental human motive is to maintain a globally positive self-evaluation. Self-affirmation theory (Steele, 1988) suggests that in case of ego threat, affirming core aspects of the self—for instance, important personal values—can help to maintain a positive self-view (for review, see Cohen & Sherman, 2014). Using self-affirmation as an intervention is beneficial in recognizing own bad habits (Harris, Mayle, Mabbott, & Napper, 2007; Harris & Napper, 2005), changing health behavior (Storr & Sparks, 2016; for a meta-analysis see Epton, Harris, Kane, van Koningsbruggen, & Sheeran, 2015), or in taking others' perspectives and being less obstinate about one's own convictions (Cohen et al., 2007; Loseman & van den Bos, 2012).

Relevant for present purposes, self-affirmation has repeatedly been found to counteract ego depletion effects (Huynh, Stefanucci, & Aspinwall, 2014; Schmeichel & Vohs, 2009; Storr & Sparks, 2016). Strength and process model again differ as to why this may be the case. Neither of the models talks about the counteracting role of self-affirmation specifically, but from the strength model's perspective, any manipulation counteracting ego depletion needs to restore or mobilize additional resources. In contrast, the process model asserts that ego depletion effects result from a loss of motivation and disengagement from the task; thus, any counteracting manipulation needs to revamp peoples' motivation for subsequent tasks to at least the baseline level prior to engaging in the highly demanding task. Thus, for a given counteracting effect of self-affirmation both models suggest different underlying mechanisms that cannot be disentangled based on the behavioral data alone. This becomes possible, however, by considering predictions on the level of exerted mental effort.

Mental Effort

Mental effort is defined as the "subjective intensification of mental and/or physical activity in the service of meeting some goal" (Inzlicht, Shenhav, & Olivola, 2018, p. 338). If a person is engaged in a task, mental effort varies with task difficulty. The more difficult a task is, the more effort an individual has to exert to complete it successfully. Importantly, mental effort is only mobilized if a person perceives engaging in a task as worthwhile (Brehm & Self, 1989; Gendolla & Richter, 2010). Accordingly, disengagement from a task is reflected in low mental effort. How can researchers measure invested mental effort? This is where psychophysiological indicators come into play—specifically, systolic blood pressure reactivity and the pre-ejection period.

Psychophysiological indicators of mental effort. People are often not fully aware of internal processes or unwilling to truthfully report them due to social desirability and self-presentational tendencies (Hofmann & Wilson, 2010; Nisbett & Wilson, 1977; van de Mortel, 2005). As a result, self-reports are limited in reliably assessing and/or reporting participants' inner mental processes, such as (a lack of) motivation and (a lack of) mental effort invested in a given activity (Brown & Bray, 2018). To circumvent these problems associated with self-reports, we used psychophysiological measures, systolic blood pressure and the pre-ejection period, as established indicators of motivation and invested mental effort.

The mobilization of effort physiologically goes along with increased activation of the sympathetic nervous system. Its antagonist, the parasympathetic nervous system, is responsible for physiological deactivation. Both systems exert influence on cardiovascular parameters (for a review, see Wright, Greenberg, & Brehm, 2003). Higher mental effort is reflected by enhanced sympathetic cardiovascular reactivity relative to a resting baseline (Richter, Gendolla, & Wright, 2016; Wright, 1996). However, the parasympathetic nervous system can mask sympathetic influences on the heart. Therefore, psychophysiological indicators of mental effort should be maximally influenced by the sympathetic and minimally influenced by the parasympathetic nervous system; they should assess sympathetic activity early in the process before parasympathetic activity has had the chance to significantly distort the measurement. Two psychophysiological indicators are presumed to fulfill the requirements of assessing sympathetic activity relatively well: systolic blood pressure (SBP) and the pre-ejection period (PEP). SBP is the maximum pressure of a pulse in the vascular system (measured in mmHg; *in*creases with increasing mental effort). In contrast, PEP reflects the time interval between the onset of the ventricular depolarization (i.e., the heart is signaled to contract) and the opening of the aortic valve—blood starts to flow (measured in milliseconds; *de*creases with increasing mental effort). Both SBP and PEP are valid and reliable indicators of sympathetic influences on the heart and often used as indicators of invested mental effort (e.g., Gendolla, Wright, & Richter, 2012; Richter & Gendolla, 2009).

Research on motivational intensity theory (Brehm & Self, 1989), in particular, has repeatedly shown that invested effort as indicated by SBP and PEP increases “as a function of task difficulty if task success is possible but drops if task success is impossible" (Richter et al., 2016, p. 157). In previous research, increasing task difficulty led to a steady increase in SBP (more mental effort), but there was no difference from the baseline when the task was impossible to master, indicating disengagement from the task (Richter, Friedrich, & Gendolla, 2008). In a similar vein, participants showed increased SBP when working on a difficult version of a counting task compared to when working on an easier version, indicating increased mental effort (easy: count forward by ones, difficult: count backward in decrements of three, Wright et al., 2007). A similar pattern emerged for PEP: Participants' PEP decreased relative to a baseline when working on increasingly difficult versions of the Sternberg task (a short-term memory task), but increased when the task was unsolvable (indicating disengagement, Richter et al., 2008).

Competing predictions from the strength versus process model. Spelling out hypotheses about mental effort for the current experimental design allows to contrast competing predictions based on the strength model and the process model. Proponents of the strength model assert that the "crucial prediction that exerting self-control reduces motivation to perform well on the second task has not been supported" (Baumeister & Vohs, 2016, p. 104). They make it very clear that "explaining all depletion phenomena based on motivation […] seems untenable" (Baumeister & Vohs, 2016, p. 105). Thus, although the model allows for exceptions, it typically predicts no difference in exerted mental effort as a function of previous self-control exertion, provided that no additional situational factor (such as self-affirmation) is present. After a previous exertion of self-control, participants should be less *able* to exert further control—this should, however, not affect task engagement, that is, the investment of mental effort in the next task.

Predictions of the process model are decisively different. From the process model's perspective, "initial acts of control lead people to become less motivated to engage in further deliberative control" (Inzlicht & Schmeichel, 2012, p. 451). Thus, mental effort should decline in the second task indicating disengagement relative to a group that faced *low demands* in the initial task (see Figure 1: Effort, A).

Adding self-affirmation to the experimental design reveals another difference in predictions between the models: According to the strength model, any intervention counteracting ego depletion effects needs to overcome the partial loss of resources in order to allow for non-impaired behavioral performance. This makes performing well on subsequent self-control tasks particularly strenuous. Thus, the model predicts that self-affirmation causes people who have previously faced high demands to invest *more* mental effort compared to those who have faced low demands in the previous task to show similar levels of self-control performance in the second task.

By contrast, the process model assumes that after facing high demands, no resource but peoples' motivation to invest further effort is reduced. Any intervention counteracting ego depletion needs to bring motivation back to baseline levels, not more. No additional effort beyond baseline levels is needed compared to participants facing low demands in the first task. Thus, no difference should emerge as a function of prior exertion of self-control in the self-affirmed condition (see Figure 1: Effort, B).

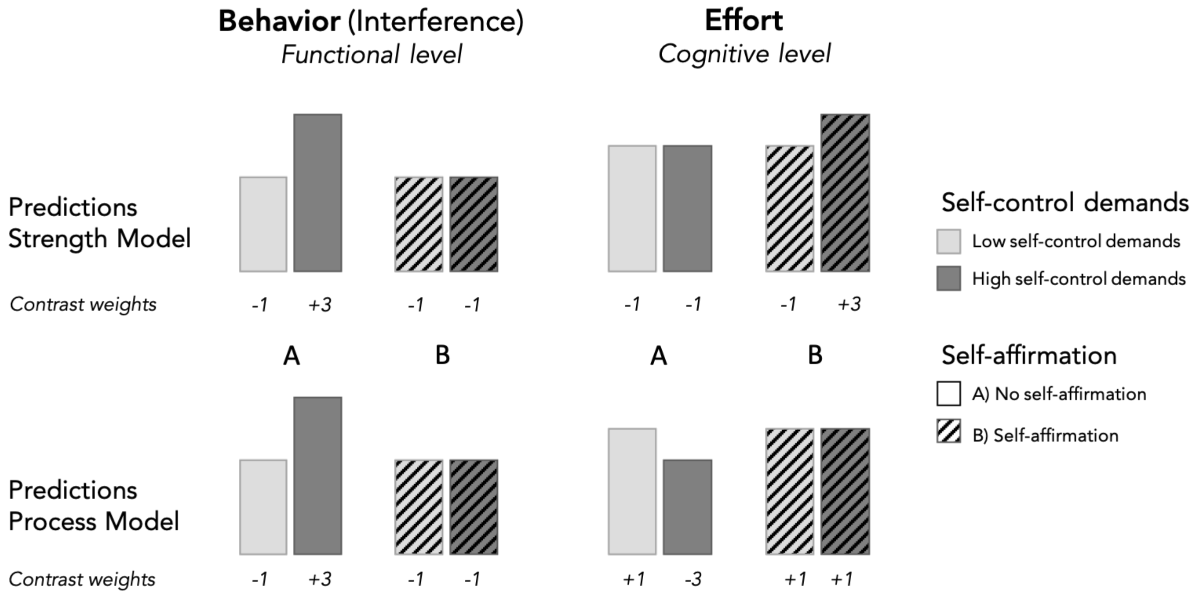


Figure 1. Predictions based on the strength model and the process model for Stroop interference (i.e., functional level) and participants' invested mental effort (i.e., cognitive level). Experimental design: Self-control demands (high vs. low) × self-affirmation (yes vs. no). Pre-registered contrast weights are shown for both the behavioral measure (Stroop interference)—for which both models predict the same outcome—and the psychophysiological indicators of mental effort with competing predictions (Furr & Rosenthal, 2003).

The Present Research

The field of ego depletion research is highly contested—and particularly mistrusted claims need particularly trustworthy evidence. To regain trust, researchers called for more theory-oriented research and transparent empirical work that follows open science principles (Friese et al., 2019; Lurquin & Miyake, 2017). In the present research, we sought to conduct a study that would make a contribution—irrespective of the specific empirical results—due to its rigorous, theory-driven model comparison approach and its stringent adherence to open science principles. Specifically, we created a context that allowed to contrast competing predictions of the strength and the process model. To this end, we added the moderator self-affirmation and the additional layer of invested mental effort as indicated by SBP and PEP to the sequential task paradigm. This design allowed us to directly compare the suitability of both models to explain the empirical pattern of results—a rare quality in ego depletion research but crucial for the advancement of the field. In addition, pre-registered studies are still infrequent in ego depletion research; the present study provides the first pre-registered conceptual replication attempt of the established moderator self-affirmation counteracting ego depletion effects.

Methods

Pre-Registration

The pre-registration to this study can be found at osf.io/cq7su. Deviations from the pre-registration are explicitly noted in the manuscript.

Materials, Data, and Code

All materials, data, and code can be found in the OSF project associated with this study at osf.io/kxb2c. All analyses were conducted in R (R Core Team, 2019); for specific packages, please see analysis script.

Reporting

We report all measures, how we determined our sample size, all data exclusions (if any), and all manipulations in the study. Non-focal and some of the pre-registered exploratory analyses are reported in the supplementary materials.

Ethical Approval

The study was approved by the local ethics committee.

Participants and Design

We tested 179 participants (M_{age} = 22.79 *SD* = 3.27; 128 females; no psychology students) in a 2 (self-control demands: high vs. low) × 2 (self-affirmation: no vs. yes) design with both factors varied between persons. Our main dependent variables were Stroop interference (based on error rates) and mental effort assessed via SBP and PEP. We restricted the sample to right-handed participants as the left hand could not be used during the study because of the psychophysiological measurement.

Sample size rationale and sequential stopping. We based our sample size calculations on the mean empirical effect of f = 0.21 for the interaction effect between self-control demands and self-affirmation found by Schmeichel and Vohs (2009; Studies 1 and 4). We registered an α = .05 and a minimum statistical power of 1-β = .80. Using these values in an a-priori power analysis with G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) resulted in a minimum sample size of N = 180 participants. In order to save resources, we deployed Bayes Factor Design Analysis (BFDA, Schönbrodt, Wagenmakers, Zehetleitner, & Perugini, 2017). This sequential procedure allows to examine the data at predefined points in the data collection process using Bayes Factors. If, at one of these points, a predefined threshold is met, data collection is stopped. Otherwise, data collection continues in pre-determined increments. We used BF_{10} = 6 as evidential threshold in favor of the alternative hypothesis and BF_{10} = 1/10 as evidential threshold in favor of the null hypothesis for the interaction effect of Stroop performance (checking for the expected pattern of the interaction). We started this process at N = 80, as pre-registered, and continued data collection until we reached the full sample as neither of the Bayes thresholds was met before.[2]

Procedure

Participants were recruited via notices and approached directly on campus. Data collection took place in single sessions with individual participants. Upon arrival at the

[2] Due to a mistake in the experimental documentation, only 179 participants were tested.

laboratory, the participant filled out an informed consent describing, among others, the electrode placement and blood pressure cuff procedures. The experimenter then attached ten pre-gelled disposable electrodes to participants' neck and torso for impedance cardiography measurement (pre-ejection period) and a blood pressure cuff on their left upper arm. The sex of the experimenter and the participant were matched to make participants feel more comfortable during the electrode placement. When all electrodes were successfully connected to the MP160 BIOPAC System (*Biopac Systems Inc.*, 2019), the recording was started and the experimenter launched the computerized tasks using the software Inquisit (*Inquisit 5*, 2016). All standardized instructions were given on the computer screen. To ensure that the experimenter was blind to the experimental conditions, the participant entered her/his ID themselves, which s/he drew randomly from an envelope containing all IDs for the current sequential-stopping phase (1-80; 81-100; 101-120; 121-140; 141-160; 161-180).

During the first ten minutes, a baseline measure for SBP and PEP (last 7 minutes, respectively) was assessed. Participants were instructed not to move and to put their feet in a box under their desk, designed to restrict excessive movement of their feet. After the baseline phase, participants were asked to indicate their current mental fatigue for a first time. Participants then practiced the Stroop task, which would later serve as the dependent variable, to prevent unwanted effects of our self-control demands manipulation on learning. The Stroop practice phase (20 trials with feedback given after every trial) was followed by the manipulation of *self-control demands (high vs. low demands)*. After a second assessment of their current mental fatigue, participants underwent the self-affirmation manipulation. Subsequently, they worked on the dependent variable and answered several questions concerning their impressions of the physiological measurement, strategies used during the experiment, and demographics. Finally, participants were debriefed, thanked, and payed for their participation.

Manipulated Variables

Manipulation of self-control demands. A counting task was used to manipulate self-control demands (see Kirschbaum, Pirke, & Hellhammer, 1993). In the *low demands* condition, participants counted upwards from zero in steps of two. Each number was entered in a new response field via the keyboard and send by clicking the 'submit'-button. The counting pace was comfortable, as each trial lasted 3,000ms before the next response field appeared. In case of blunder, participants were asked to continue as if nothing had happened. In the *high demands* condition, participants were asked to count backwards in steps of 13 as fast as possible starting from a random number between 4,020 and 4,820. There was no time limit to account for individual differences in participants' mathematical aptitude. In case of blunder, participants were required to restart from a new random number. In both demand conditions, the counting task lasted 8 minutes.

Self-affirmation manipulation. Participants in the self-affirmation condition read a list of eleven values and ranked these values by their personal importance. Participants then wrote an essay explaining why their highest ranked value is important to them and to describe a time in their lives when this value had been particularly important (Cohen, Aronson, & Steele, 2000). This manipulation is one of the most often used self-affirmation manipulations (McQueen & Klein, 2006) and has repeatedly been used successfully as a manipulation to counteract ego depletion (see Schmeichel & Vohs, 2009; Storr & Sparks, 2016). In the control condition, participants were instructed to write an essay describing an ordinary, grey hole puncher placed on the desk in from of them in as much detail as possible (used similarly in previous self-affirmation research; e.g., Miketta & Friese, 2019). Describing neutral objects instead of writing about personally less important values has the benefit that it constitutes a similar kind of activity as in the self-affirmation condition, but is not concerned with values in any ways which could have unwanted downstream effects on the following task. Both writing tasks lasted 6 minutes.

Measured Variables

Stroop task. A color Stroop task served as the behavioral dependent variable (Luethi et al., 2016). The words "red", "yellow", and "blue" were presented in three different font colors. Participants had to indicate the font color of the word as quickly as possible, while ignoring the literal meaning. The font color either matched (congruent trial) or did not match (incongruent trial) the literal meaning of the word. Each trial consisted of the presentation of a word for 1,500ms on a black screen followed by an inter-stimulus interval of 500ms. If a participant did not press a key during these 2,000ms a warning message appeared asking for a quicker reaction.

We created a modified version of the Stroop task. The word "yellow" served as an exception: Whenever "yellow" appeared, participants had to indicate the literal meaning of the word instead of its font color. This modification made sure that participants had to look at the center of the screen and actually read each stimulus instead of looking at the periphery to detect the color of the stimuli without actually reading them, thereby circumventing the self-control demanding characteristics of the task. Each participant completed 4 blocks of 45 trials each (56% congruent, 28% incongruent, 16% "yellow" exception rule) with no breaks between the blocks. The order of trials was pseudorandomized and fixed.

Psychophysiological measures. Each psychophysiological indicator was measured during four experimental periods: Baseline, manipulation of demands, self-affirmation manipulation, and Stroop task.

Systolic blood pressure. We used a *GE Carescape Dinamap V100* to measure participants' SBP every 60sec during the relevant experimental periods. For the baseline, we calculated the mean of all measures (last 7 minutes). For the remaining tasks, we calculated the reactivity in SBP during the respective task relative to the baseline (difference between mean SBP during the task minus mean SBP during the baseline). We measured diastolic blood pressure and heart rate for exploratory reasons. Reactivity for the latter two measures was calculated in the same manner.

Pre-ejection period. We used an *ECG100C* module to obtain the ECG signal and a *NICO100C* module to obtain the cardiac impedance signal dZ/dt both wired to a *MP150* polygraph (Biopac Systems Inc., Goleta, CA, USA). The sampling rate was 1 kHz. We equipped each participant with 10 spot electrodes. First, participants' skin was abraded for optimal contact and additional gel was applied to the electrodes. Next, the upper two voltage electrodes were placed on the left and right side of the base of the neck, the lower two on the right and left side of the torso at the level of the lower end of the sternum. The current electrodes were placed at a distance of 3cm from the respective voltage electrodes. The remaining two electrodes for the ECG signal were placed below the right collarbone and on the lowest left rib. Finally, we connected the leads, put them in loops and fastened them with adhesive tape for optimal signal quality (see Figure 2).

We used the software *AcqKnowledge 4.3* (Biopac Systems Inc., Galeta, CA) to record and process the psychophysiological signal. Pre-ejection period for each of the relevant periods was determined and calculated using partially automated scripts. To detect the B point, we used the third derivative classification with the local maximum found within 300ms to the C point. As there are several possible algorithms to detect the B point, we used the method suggested as superior by recent research examining the efficacy of different detection algorithms (Árbol et al., 2017). The penultimate entire cycle of the baseline served as representative cycle for the automated analysis. For all tasks, we used PEP reactivity relative to the baseline level.

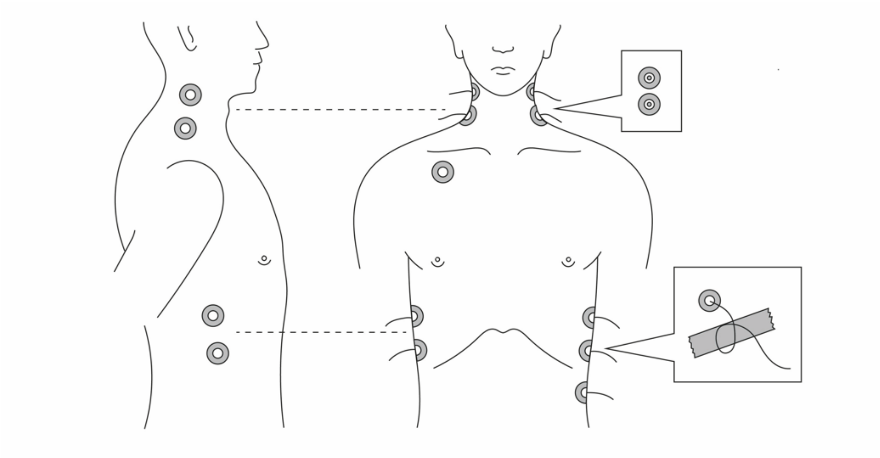


Figure 2. Spot electrode placement with voltage electrodes located at the base of the neck and at the level of the lower end of the sternum. Current electrodes were placed at a distance of 3cm from the respective voltage electrodes. ECG electrodes were placed below the right collarbone and on the lowest left rib. For optimal quality, the leads were put in loops and fastened with adhesive tape.

Mental fatigue. Current mental fatigue was assessed with three items of the German version of the State Self-Control Capacity Scale (SSCCS; items 1, 12, & 15; Bertrams & Dickhäuser, 2009) and two additional items (“I feel mentally drained”, “I feel mentally efficient”) rated on a 7-point-scale (1 = *strongly disagree* to 7 = *strongly agree*). Current mental fatigue was assessed both at the beginning of the study and after the manipulation of self-control demands (α = .89, .87).

Manipulation check counting task (manipulation of self-control demands). After the counting task, participants were asked, “How hard did you have to concentrate during the counting task?” and “How exhausting was it to work on the just finished counting task?”.

Participants were asked to compare the task to a math exam of 30 minutes. Answers were given on 7-point-scales (1 = *much less than in an exam* to 7 = *just like in a math exam*; $\rho = .83$[3]).

Morningness/Eveningness. We exploratorily assessed morningness versus eveningness using the last item of the D-MEQ (D-MEQ; item 19; Griefahn, Künemund, Bröde, & Mehnert, 2001).

Demographic and control questions. Final questions assessed demographic information, self-rated data quality, strategies during the Stroop task, presumed study purpose, and subjective perceptions of the psychophysiological measurement. The full list of items and their English translations can be found in the associated OSF project.

Results

Manipulation Check (Pre-Registered)

As expected, participants in the *high demands* conditions felt that the counting task required more concentration and was more exhausting than participants in the *low demands* conditions ($t_{\text{one-sided}}[176.55] = -11.32$, $p < .001$, $d = -1.69$, 95% CI [$-\infty$, -1.41], $M_{\text{high demands}} = 4.90$, $SD_{\text{high demands}} = 1.35$, $M_{\text{low demands}} = 2.64$, $SD_{\text{low demands}} = 1.33$).

Behavioral Self-control Performance based on Stroop Interference

We pre-registered two different outlier-cutoffs for the behavioral Stroop data. The primary, confirmative criterion—± 2.5 *SD*s from the respective mean—is one that we have used in previous research (e.g., osf.io/my5x2). The secondary, exploratory criterion was ± 3 median absolute deviations (MAD, Leys, Delacre, Mora, Lakens, & Ley, 2019; Leys, Ley, Klein,

---

[3] Spearman-Brown reliability coefficient, recommended for two item scales (Eisinga, Grotenhuis, & Pelzer, 2013).

Bernard, & Licata, 2013). The latter is sometimes preferred because MADs are based on a measure of central tendency—the median—that is not influenced by the very outliers that it is supposed to detect. We report results based on both the 2.5 *SD* criterion and the 3 MAD criterion.

As pre-registered, we excluded Stroop trials with reaction times < 200ms, individuals whose Stroop interference deviated more than the respective outlier-cutoff from the condition mean, and individuals whose total number of errors on congruent trails deviated more than the respective outlier-cutoff from the overall mean.

Outlier-cutoff 2.5 SDs (pre-registered). In line with the predictions of both the strength and the process model (see Figure 1: Behavior, A), participants in the *non-affirmed* conditions performed worse in the Stroop task after exerting self-control than after not exerting self-control. However, this difference was not significant ($t_{\text{one-sided}}[80.49] = -1.59$, $p = .058$, $d = -0.35$, 95% CI [-0.79, 0.09]; for descriptive statistics, see Table 1).

Contrary to predictions of both models, the self-affirmation manipulation did not counteract the detrimental effect of the ego depletion manipulation (pre-registered contrast: *high demand+non-affirmed* = +3; remaining three conditions = -1; Furr & Rosenthal, 2003; see Figure 1: Interference). Quite in contrast, participants in the *self-affirmed* conditions showed similarly poor Stroop performance as participants in the *high demand+non-affirmed* condition. Our pre-registered contrast thus did not fit the data ($F[1, 168] = 0.39$, $p = .536$, $f = 0.05$; see Figure 2; for descriptive statistics see Table 1).

Table 1

Stroop interference on errors (outlier cutoff 2.5 SDs from condition means; % errors on incongruent trials minus % errors on congruent trials).

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	10.38 (6.70) *N* = 41	10.09 (7.47) *N* = 45	*10.23 (7.07)*
	Low	8.15 (6.13) *N* = 43	10.61 (7.11) *N* = 41	*9.37 (6.71)*
	Overall	*9.24 (6.47)*	*10.34 (7.26)*	

Note. M (SD).

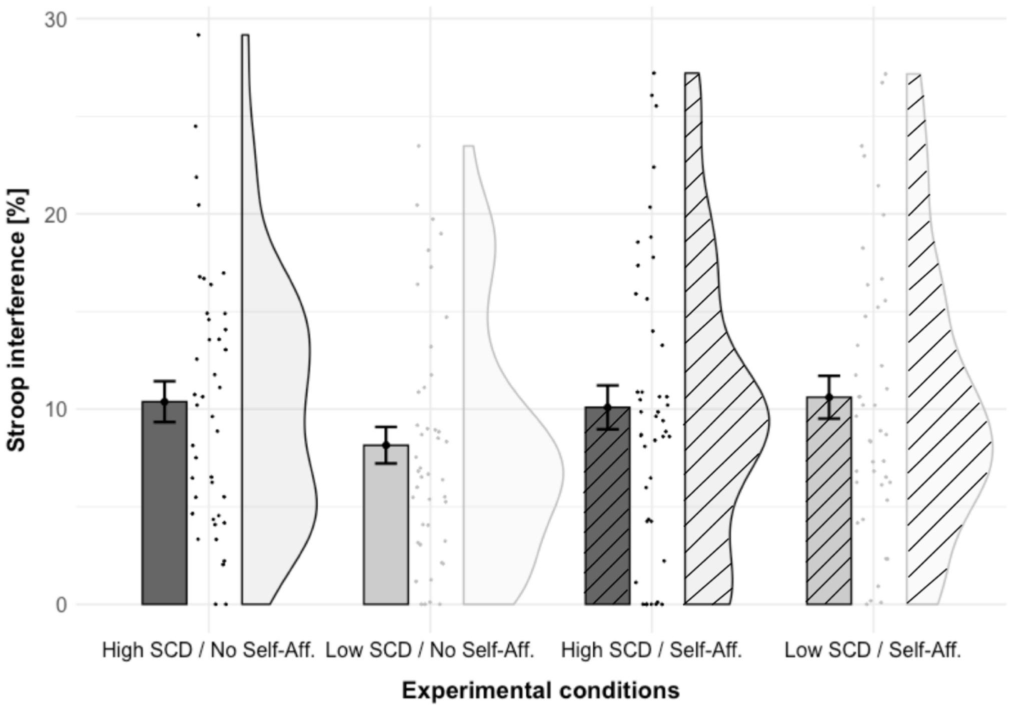


Figure 3. Stroop task: Interference on errors (% errors on incongruent trials - % errors on congruent trials) as a function of self-control demands in the first task and self-affirmation condition.

Bayesian model comparisons (exploratory)[4]**.** Comparing our pre-registered model (*high demand+non-affirmed* > *remaining three conditions equal*) against a model with no differences in Stroop performance ("all-equal") resulted in a BF = 4.18 in favor of the "all-equal" model. The best fit was provided by the model specified as *low demand+non-affirmed* < *remaining three conditions equal,* both compared to the pre-registered model (BF = 5.57), and the "all-equal model", although this model showed almost equal fit (BF = 1.33). Thus, the data most supported the model stating that participants in the *low demand+non-affirmed* condition performed better than participants in the remaining three conditions who all showed equally impaired performance.

Outlier-cutoff 3 MADs (pre-registered, exploratory). As predicted by both the strength model and the process model, participants in the *non-affirmed* conditions performed worse in the *high demand* condition than in the *low demand* condition (t[79.63] = -1.76, p = .041, d = -0.39, 95%CI[-∞, -0.02]). Contrary to our standard outlier cutoff of more than ±2.5 *SD*, this difference was statistically significant based on our pre-registered significance level of $p < .05$.

Compared to the standard outlier-cutoff, the results do not change with regard to the moderating role of self-affirmation. Self-affirmation did not counteract the detrimental effects of high self-control demands. To the contrary, participants in both *self-affirmed* conditions showed similar performance as participants in the *high demands+non-affirmed* condition. Again, our pre-registered contrast did not fit the data (F[1, 168] = 0.77, p = .381, f = 0.07; for descriptive statistics see Table 2).

[4] A Bayes Factor of BF = X indicates that the hypothesis is X-times more likely than the hypothesis compared to given the observed data (for an introduction, see Hoijtink, Mulder, van Lissa, & Gu, 2019). All Bayes Factors were obtained using informative hypothesis testing with the *bain* package in *R* using default settings (see Hoijtink et al., 2019).

Bayesian model comparisons (exploratory). As for our standard outlier cutoff, the best fit was again provided by the model specified as *low demands+non-affirmed < remaining three conditions equal*, both relative to the pre-registered contrast (BF = 4.76) and to the "all-equal" model (BF = 1.53)[5]. Again, our data thus provide most support for the model assuming that participants in the *low demand+non-affirmed* condition performed better than participants the remaining three conditions, all showing similarly impaired performance.

Table 2

Stroop interference based on errors (outlier cutoff ±3 MADs; % errors on incongruent trials - % errors on congruent trials)

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	10.79 (6.95) N = 41	10.79 (7.96) N = 45	*10.79 (7.45)*
	Low	8.26 (6.13) N = 43	10.03 (6.10) N = 41	*9.12 (6.15)*
	Overall	*9.49 (6.63)*	*10.43 (7.11)*	

Note. M (SD).

Interim summary. As expected, participants in the *non-affirmed* conditions showed an impaired performance in the *high demand* compared to the *low demand* condition. However, participants in the *self-affirmed* conditions did not perform as well as expected. In contrast to our expectations and in contrast to previous research (Huynh et al., 2014; Schmeichel & Vohs, 2009), self-affirmation did not counteract the detrimental effect of a demanding first task.

[5] Comparing the pre-registered contrast against the "all-equal" model resulted in BF = 3.11 in favor of the latter.

Mental Effort

The predictions of the strength and the process model differ markedly concerning participants' investment of mental effort. We contrasted the predictions of both models on mental effort as indicated by SBP and PEP. As pre-registered, we excluded participants deviating more than 2.5 *SD*s from the respective condition mean. Using our exploratory outlier cutoff of 3 MADs did not meaningfully change the results (see supplementary materials).

Systolic blood pressure (pre-registered). As expected, participants showed no difference in SBP during the baseline (2 × 2 ANOVA with *self-control demands* × s*elf-affirmation*: all *p*s > .125). Also, as expected, participants in the *high* compared to the *low-demands* conditions exerted more mental effort during the manipulation of self-control demands—difficult versus easy counting task—as indicated by higher SBP relative to the baseline (t[165.38] = -4.65, $p < .001$, d = -0.71, 95%CI [-∞, -0.45]). This speaks to the strength of our manipulation in terms of required mental effort and suggests that the SBP measurement was successful in picking up participants' increased effort in the *high* versus the *low demand* conditions.

During the self-affirmation manipulation, the four experimental conditions did not differ in invested mental effort (*self-control demands* × *self-affirmation* ANOVA: all *p*s > .158). Also, there were no differences between conditions as indicated by SBP reactivity during the Stroop task—neither as a function of self-control demands, nor as a function of self-affirmation or their interaction (all *p*s > .259; see Figure 4; for descriptive statistics see Table 3).

Bayesian model comparisons (exploratory). We used Bayesian informed hypothesis testing to compare the predictions of the strength and the process model (see Figure 1: Effort) with the "all-equal" model and the model specified as *low demands+non-affirmed* < *remaining* *three conditions equal.* The best fit was indicated for the "all-equal" model (all BFs > 10). The "*low demands+non-affirmed < remaining three conditions equal"* model provided no better fit to the data than the models derived from the predictions of the strength model (BF = 1.12) and the process model (BF = 1.34; see Table 3).

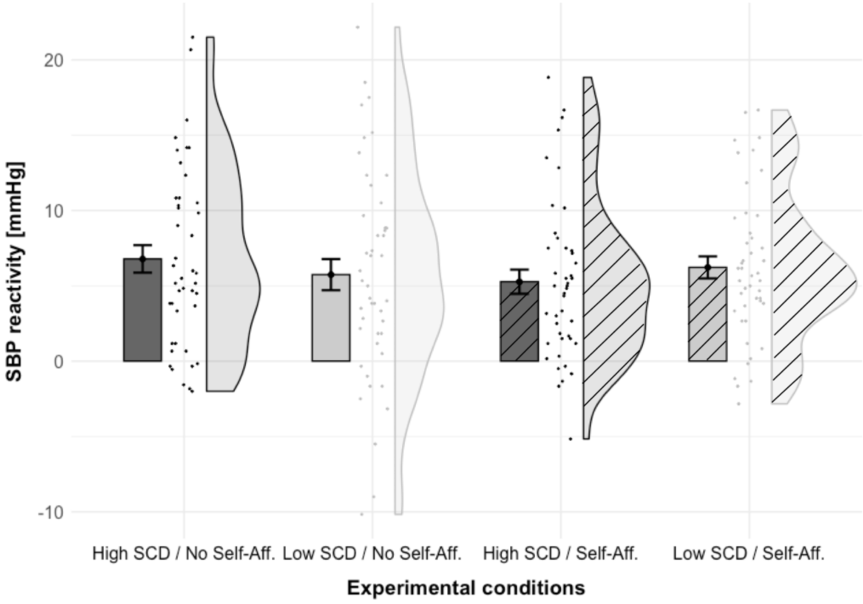


Figure 4. Systolic blood pressure reactivity during the Stroop task as a function of *self-control demands* and *self-affirmation*. Higher bars indicate increased invested mental effort.

Table 3

Systolic blood pressure (SBP)

1) Systolic blood pressure (SBP) in mmHg

A) Mean SBP during baseline.

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	113.14 (9.21) *N* = 42	110.74 (7.56) *N* = 45	*111.90 (8.43)*
	Low	110.28 (7.94) *N* = 45	111.96 (10.27) *N* = 45	*111.12 (9.17)*
	Overall	*111.66 (8.64)*	*111.35 (8.99)*	

B) Reactivity: Manipulation of self-control demands.

Self-control demands	High	6.54 (5.27) *N* = 86
	Low	3.13 (4.39) *N* = 89

C) Reactivity: Self-affirmation manipulation.

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	6.50 (6.95) *N* = 43	6.09 (3.90) *N* = 44	*6.29 (5.59)*
	Low	4.28 (6.06) *N* = 45	6.24 (4.88) *N* = 43	*5.24 (5.57)*
	Overall	*5.36 (6.57)*	*6.16 (4.39)*	

D) Reactivity: Stroop task.

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	6.92 (6.00) *N* = 43	5.61 (5.88) *N* = 44	*6.25 (5.94)*
	Low	5.76 (7.00) *N* = 46	6.22 (4.90) *N* = 45	*5.99 (6.03)*
	Overall	*6.31 (6.35)*	*5.92 (5.39)*	

2) Bayesian model comparisons for invested mental effort as indicated by SBP (reactivity)

	Process model over …	*low demands+non-affirmed* < *remaining 3 conditions equal* over …	All equal over …
Strength model	BF = 0.83	BF = 1.12	BF = 12.19
Process model		BF = 1.34	BF = 14.06
low demands+non-affirmed< *remaining 3 conditions equal*			BF = 10.88

Note. M (SD).

Pre-ejection period (pre-registered). As expected, participants showed no difference in PEP during the baseline (ANOVA *self-control demands* × *self-affirmation*: all *p*s > .442; for descriptive statistics see Table 4). During the manipulation of self-control demands, participants in both conditions invested more mental effort as indicated by shortened PEP compared to the baseline ($t_{\text{one sample}}[146] = -2.03$, $p = .045$, $d = -0.17$, 95%CI [-.033, 0.00]). However, we did not find the expected difference in invested mental effort as a function of self-control demands ($d = -0.11$, 95% CI [-0.44, 0.22], $M_{\text{high demands}} = -0.78$, $SD_{\text{high demands}} = 6.80$, $M_{\text{low demands}} = -1.54$, $SD_{\text{low demands}} = 7.18$). Thus, the PEP measurement did not indicate the increased mental effort that was necessary to work on the counting task in the *high* compared to the *low demand* condition and that was indicated by both the questionnaire-based manipulation check and the SBP measurement. We therefore stopped the analysis of PEP data at this point as there was no evidence that the PEP measure picked up the differential levels of mental effort exerted by participants in the *high* versus *low demand* conditions (see SBP reactivity and self-report measures).

Table 4

Pre-ejection period (PEP) in milliseconds

A) Mean PEP during baseline.				
		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	109.88 (13.22) *N* = 38	111.76 (14.37) *N* = 41	*110.86 (13.77)*
	Low	113.17 (17.14) *N* = 42	111.32 (15.22) *N* = 36	*112.31 (16.21)*
	Overall	*111.61 (15.40)*	*111.55 (14.68)*	
B) Reactivity: Manipulation of self-control demands.				
Self-control demands	High	-0.78 (6.80) *N* = 75		
	Low	-1.54 (7.18) *N* = 72		

Note. M (SD).

SBP progression over the course of the study (exploratory). All pre-registered analyses relating to mental effort referred to reactivity (i.e., change) relative to the baseline measurement at the beginning of the study. On an exploratory basis, we also examined SBP progression over the course of the study, that is, relative to the previous task in the experimental procedure (see Figure 5; for descriptive statistics see Table 5).

SBP progression: Self-affirmation manipulation. Relative to the previous task (self-control demands manipulation), participants showed a stronger increase in mental effort while affirming core personal values compared to describing the hole puncher (see the pronounced upward progression for dotted lines in Figure 4 relative to the self-control demands manipulation; main effect self-affirmation: $F[1, 175] = 12.08$, $p = .001$, $\eta_p^2 = 0.07$; for descriptive statistics see Table 5A). This increase in SBP was also particularly pronounced for participants who did the *low demands* task beforehand (main effect self-control demands: $F[1, 175] = 19.38$, $p < .001$, $\eta_p^2 = 0.09$). This latter finding may reflect that it was easier for participants who initially did not exert a lot of effort (i.e., *low demands* conditions) to increase their mental effort than it was for participants who exerted a lot of effort in the *high* demands conditions. The interaction between *self-control demands* and *self-affirmation* was not significant ($F[1, 175] = 0.15$, $p = .699$, $\eta_p^2 < 0.01$).

SBP progression: Stroop task. During the Stroop task and relative to the self-affirmation manipulation, participants in the s*elf-affirmed* conditions showed a decrease in invested mental effort—contrary to participants in the *non-affirmed* conditions (see decline for dotted lines in Figure 4 from self-affirmation to Stroop phase; main effect self-affirmation: $F[1, 175] = 4.97$, $p = .027$, $\eta_p^2 = 0.03$; for descriptive statistics see Table 5B). Only participants in the *low demands+non-affirmed* condition showed a pronounced increase in invested mental effort during the Stroop task relative to the previous task (i.e., incline for continuous grey line). As these were the only participants who showed superior performance during the Stroop task, these findings provide a possible explanation for the unexpected behavioral effect of the moderator self-affirmation (i.e., participants in the s*elf-affirmed* conditions performed poorer, not better, than those in the *low demands+non-affirmed* condition): Participants in the *self-affirmed* conditions showed a stronger increase in invested mental effort during the self-affirmation manipulation relative to the previous task than those in the *non-affirmed* conditions. We did not expect the self-affirmation manipulation to evoke a more pronounced investment in mental effort relative to the control task (describing a hole puncher). However, the additional invested mental effort during the affirmation of core personal values relative to the previous task could explain why participants subsequently invested relatively less mental effort during the Stroop task—and therefore why they did not differ from participants in the *high demand+non-affirmed* condition.

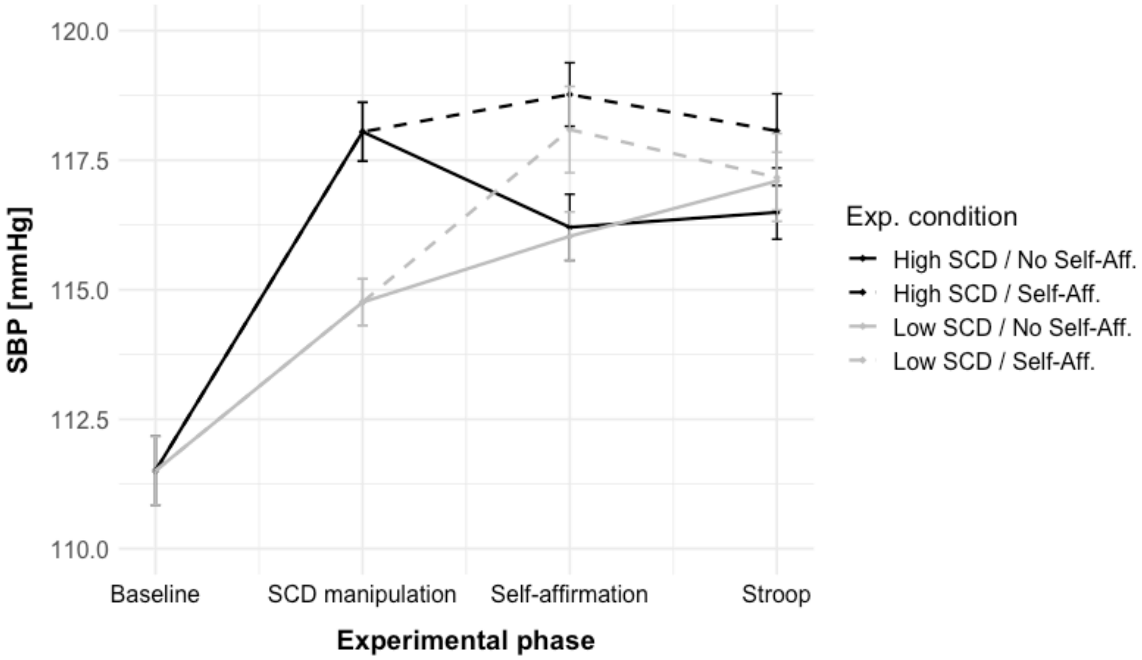


Figure 5. SBP progression relative to the previous task$_{X-1}$ over the course of the experiment. The more pronounced increase from Baseline to the self-control demands manipulation for the two *high demand* conditions (black) versus the *low demands* conditions (grey) reflects the traditional SBP reactivity (relative to Baseline). SBP progression from thereon indicates a markedly more pronounced reaction in the *self-affirmed* conditions (dashed lines) compared to the *non-affirmed* conditions (continuous lines). During the Stroop task, only participants in the *low demands+non-affirmed* condition (continuous grey line) invested more mental relative to the previous task.

Table 5

SBP progression over the course of the study

A) SBP progression: Self-affirmation manipulation relative to self-control demands manipulation				
Descriptive statistics		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	-1.84 (4.18) *N* = 43	0.72 (4.11) *N* = 45	*-0.53 (4.32)*
	Low	1.27 (3.19) *N* = 46	3.33 (5.59) *N* = 45	*2.29 (4.63)*
	Overall	*-0.23 (4.00)*	*2.03 (5.05)*	
B) SBP progression: Stroop relative to self-affirmation manipulation				
Descriptive statistics		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	0.29 (3.39) *N* = 43	-0.70 (4.80) *N* = 45	*-0.22 (4.18)*
	Low	1.07 (3.78) *N* = 46	-0.92 (5.69) *N* = 45	*0.08 (4.89)*
	Overall	*0.69 (3.60)*	*-0.81 (5.24)*	

Note. *M (SD).* No outliers excluded.

Discussion

The main aim of this study was to contrast competing predictions from two theoretical accounts that try to explain ego depletion effects in psychologically very different ways: the strength model (Baumeister & Vohs, 2016) and the process model (Inzlicht & Schmeichel, 2012). On the functional level of analysis in terms of observable behavior (De Houwer, 2011), both models predict poorer self-control performance after the initial exertion of self-control. On the cognitive level of analysis concerning the underlying psychological mechanisms, the models provide markedly different explanations for the same behavioral phenomenon. To contrast these competing predictions, we combined the traditional sequential task paradigm with a moderator approach and with psychophysiological indicators of mental effort. Together, this design allows to test competing predictions of the two models.

Disentangling Two Theoretical Approaches

The prerequisite to contrast the predictions of the two models concerning underlying mechanisms as planned was that we would replicate an established moderator effect—that is, self-affirmation should counteract ego depletion on the behavioral level. Despite abundant evidence suggesting otherwise (Huynh et al., 2014; Schmeichel & Vohs, 2009; Storr & Sparks, 2016), the affirmation of core personal values had the opposite effect. Although we found moderate evidence supporting the traditional ego depletion effect in the *non-affirmed* conditions, *self-affirmed* participants who had previously worked on a demanding counting task did not show improved self-control performance during the Stroop task. To the contrary, performance in the two self-affirmed conditions was similarly poor as that of participants who had engaged in a demanding task without subsequent self-affirmation.

Even if—due to the unexpected effect of the moderator—we could not test the competing patterns for all four experimental conditions, our results for invested mental effort and self-control behavior provide preliminary evidence as to the fit of the two models, particularly for the *non-affirmed* conditions. First of all, the robust effect of the self-control demands manipulation on SBP reactivity as well as on self-reports suggests that SBP was a suitable measure of mental effort. PEP reactivity did not capture the more pronounced mental effort investment for participants in the *high* versus *low demand* conditions. We therefore refrained from interpreting PEP measures any further.

Concerning the emerging pattern in the Stroop task for participants in the *non-affirmed* conditions, our data on invested mental effort as indicated by SBP reactivity provide less support for the process model than for the strength model: Participants in the *high demands+non-affirmed* condition did not invest less mental effort than participants in the *low demands+non-affirmed* condition (see exploratory Bayesian model comparisons). This is in line with the predictions of the strength model but not with the process model that predicts a decrease in invested mental effort due to decreased motivation after the *highly demanding* task without *self-affirmation*.

Unexpected Effect of the Self-Affirmation Manipulation

We expected the self-affirmation manipulation to counteract a behavioral ego depletion effect and lead to similar behavioral performance as for participants who initially engaged in a hardly demanding task. The opposite effect occurred: After self-affirmation, participants performed poorly on the Stroop task. There are several possible explanations for this unexpected finding. First, participants not affirming core personal values described a hole puncher in great detail. They thus focused on a specific object during several minutes with a narrow focus of attention. This narrow focus of attention might have been beneficial for the performance in the subsequent Stroop task that requires a focused, narrow mindset instead of a global, abstract mindset. In previous studies on the counteracting role of self-affirmation on the ego depletion effect, the affirmation of core personal values was beneficial for tasks as the cold pressor challenge and delay of gratification (Schmeichel & Vohs, 2009). Potentially, value affirmation may be beneficial in these cases, as pain tolerance and short-term sacrifice in favor of future gains might profit more from the abstract and global “big picture” that the self-affirmation manipulation may have evoked, rather than the focus on details (e.g. how much it hurts and how long it takes). In a similar way, affirming core personal values, broadening the focus and thinking of the “big picture” instead may have impaired participants performance in the Stroop task. This would be in line with previous research suggesting higher construal level as a mechanism potentially underlying the counteracting effect of self-affirmation on ego depletion phenomena (see Schmeichel & Vohs, 2009). Corroborating our reasoning, recent research argues that people are able to distinguish between tasks that profit from *low-level* versus *high-level* construal and to prepare themselves accordingly to promote performance (Nguyen, Carnevale, Scholer, Miele, & Fujita, 2019). Most relevant to present purposes, this work suggests that people understand that a concrete, low level mindset is beneficial for performance on a Stroop task.

Second, compared to their most important values in life as family and friends, participants may have felt that the task was hardly worth the effort. Thus, disengagement due to perceived low importance of success could explain their performance. If this was the case, the results could be interpreted in favor of the process model: As participants were no longer motivated, they showed worse performance in the Stroop task. In sum, these results pave the way for a new and testable hypothesis concerning the counteracting effect of self-affirmation on ego depletion effects: Self-affirmation may be beneficial for performance on subsequent self-control tasks that profit from an abstract mindset (such as making decisions about the future in a delay-of-gratification task). It may be detrimental, however, for tasks that profit from a narrow focus of attention—such as the Stroop task. Future research should investigate this hypothesis by experimentally varying the dependent variable. Such work would also contribute to the discussion about the suitability of various different self-control related dependent variables in ego depletion research and their often low correspondence (Duckworth & Kern, 2011; Friese et al., 2019).

Second, our exploratory results on invested mental effort as indicated by SBP progression over the course of the study suggest that participants invested more mental effort during the affirmation of core personal values and then less in the following Stroop task. Thus, as the value affirmation was more mentally demanding than the control task and as *self-affirmed* participants then invested less mental effort during the Stroop task, it may not be surprising that their performance was impaired and comparable to those of participants in the *high demands+non-affirmed* condition. This speculation is hinted to by the data and post hoc analyses of SBP progression. Future research may systematically investigate this possibility.

Recently, researchers from the field of behavioral intensity who study determinants and cardiovascular correlates of effort provided an intriguing “outsider’s perspective” on ego depletion research. They proposed two main issues that may have contributed to its deficient replicability: a) the difficulty to consistently evoke ego depletion effects even with highly standardized protocols and b) the possibility that the effect may only emerge in a behavioral-restraint “sweet spot” (Wright, Mlynski, & Carbajal, 2019). Although a thorough analysis of the perspective is beyond the scope of this article, the notion of the *sweet spot* is particularly relevant for the present study. Specifically, the sweet spot—“a multidimensional motivational space wherein rested study participants view restraint as possible and worthwhile and fatigued study participants do not” (Wright et al., 2019, p. 6)—provides an explanation similar to our interpretation of the present findings. Participants in the *self-affirmed* conditions may not have seen restraint in the Stroop task as worthwhile. Thus, even if they may have been able (i.e., sufficiently rested) to master the Stroop task’s difficulty, a lack of success importance may have reduced the amount of effort invested and deteriorated task performance.

Limitations

There are several limitations to our study. First, the behavioral ego depletion effect was not significant based on the outlier cutoff of 2.5 standard deviations, although the descriptive result was as expected and the effect size was small-to-moderate (p = .058, d = -0.35). However, the effect did emerge significant as predicted when excluding outliers based on the pre-registered exploratory outlier criterion using median absolute deviations. This criterion is less influenced by extreme values and is therefore recommended by some researchers over outlier criteria based on standard deviations from the mean (Leys et al., 2019, 2013).

Second, our psychophysiological indicators of invested mental effort showed mixed results in terms of convergent validity. As expected, neither of the two measures indicated differences in invested mental effort during the baseline. During the manipulation of self-control demands, however, only the SBP pattern matched the hypothesis and the self-reports for invested mental effort: Participants in the *high demand* conditions invested more mental effort than participants in the *low demand* conditions. By contrast, for PEP there were no differences between conditions as a function of self-control demands. As the SBP findings and the self-reports speak to the strength of the self-control demands manipulation, we would have expected differences in PEP as well. We therefore did not analyze PEP reactivity any further.

Our exploratory results on SBP progression relative to the preceding task as opposed to relative to the baseline suggest that this comparison in terms of changes in invested mental effort during the course of a study may be a promising one. The emerging pattern using this progression measure indicated that participants showed a more pronounced increase in invested mental effort while affirming core personal values and less so during the self-affirmation period after the highly self-control demanding task. This measure also indicated that participants in the *low demands+non-affirmed* condition showed the strongest increase in invested mental effort during the Stroop task which fits the performance pattern in the Stroop task as they also showed the best performance. Future research could explore whether and when using psychophysiological progression relative to the preceding task constitutes an appropriate measure.

Conclusion

We sought to contrast the predictions of the strength and the process model regarding possible underlying mechanisms that cause ego depletion effects. To this end, we combined the traditional sequential task paradigm with a) the moderator self-affirmation and b) established psychophysiological indicators of mental effort. In principle, this combination allows to close a gap: The approach enables researchers to contrast competing predictions to promote theory development. As the self-affirmation manipulation had an unexpected, negative effect on self-control performance, the pre-registered replication of this moderator was not successful and we were thus limited in our analysis contrasting the models' competing predictions. Future research investigating self-affirmation manipulations might reflect upon the moderator's specific effect on different kinds of tasks with a focus on mental construal levels. In addition, the progression of invested mental effort over the course of time might be a promising avenue for future research.

Acknowledgments

We thank Annika Scheuß, Jan Helge Kaben, Nicholas Jakobson and Niklas Klever for their help and perseverance with data collection.

References

Árbol, J. R., Perakakis, P., Garrido, A., Mata, J. L., Fernández-Santaella, M. C., & Vila, J. (2017). Mathematical detection of aortic valve opening (B point) in impedance cardiography: A comparison of three popular algorithms. *Psychophysiology*, *54*, 350–357.

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.

Baumeister, R. F., & Vohs, K. D. (2016). Strength model of self-regulation as limited resource: Assessment, controversies, update. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (Vol. 54, pp. 67–127). San Diego, CA, US: Academic Press.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, *16*, 351–355.

Bertrams, A., & Dickhäuser, O. (2009). Messung dispositioneller Selbstkontroll-Kapazität: Eine deutsche Adaptation der Kurzform der Self-Control Scale (SCS-K-D). *Diagnostica*, *55*, 2–10.

Biopac Systems Inc. (2019). Goleta, CA, USA.

Brehm, J. W., & Self, E. A. (1989). The Intensity of Motivation. *Annual Review of Psychology*, *40*, 109–131.

Brown, D. M. Y., & Bray, S. R. (2018). Effects of mental fatigue on exercise intentions and behavior. *Annals of Behavioral Medicine*. https://doi.org/10.1093/abm/kay052

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.

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.

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, *26*, 1151–1164.

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology*, *65*, 333–371.

Cohen, G. L., Sherman, D. K., Bastardi, A., Hsu, L., McGoey, M., & Ross, L. (2007). Bridging the partisan divide: Self-affirmation reduces ideological closed-mindedness and inflexibility in negotiation. *Journal of Personality and Social Psychology*, *93*, 415–430.

Dang, J., Liu, Y., Liu, X., & Mao, L. (2017). The ego could be depleted, providing initial exertion is depleting. *Social Psychology*, *48*, 242–245.

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.

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.

Eisinga, R., Grotenhuis, M. te, & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637–642.

Epton, T., Harris, P. R., Kane, R., van Koningsbruggen, G. M., & Sheeran, P. (2015). The impact of self-affirmation on health-behavior change: A meta-analysis. *Health Psychology*, *34*, 187–196.

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.

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.

Faul, F., Erdfelder, E., Lang, 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.

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.

Furr, R. M., & Rosenthal, R. (2003). Evaluating theories efficiently: The nuts and bolts of contrast analysis. *Understanding Statistics*, *2*, 33–67.

Garrison, K. E., Finley, A. J., & Schmeichel, B. J. (2019). Ego depletion reduces attention control: Evidence from two high-powered preregistered experiments. *Personality and Social Psychology Bulletin*, *45*, 728–739.

Gendolla, G. H. E., & Richter, M. (2010). Effort Mobilization when the Self is Involved: Some Lessons from the Cardiovascular System. *Review of General Psychology*, *14*, 212–226.

Gendolla, G. H. E., Wright, R. A., & Richter, M. (2012). Effort intensity: Some insights from the cardiovascular system. In *The Oxford Handbook of Human Motivation* (pp. 420–438). Oxford: Oxford University Press.

Griefahn, B., Künemund, C., Bröde, P., & Mehnert, P. (2001). Zur Validität der deutschen Übersetzung des Morningness-Eveningsness-Questionnaires von Horne und Östberg. *Somnologie*, *5*, 71–80.

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.

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.

Harris, P. R., Mayle, K., Mabbott, L., & Napper, L. (2007). Self-affirmation reduces smokers' defensiveness to graphic on-pack cigarette warning labels. *Health Psychology*, *26*, 437–446.

Harris, P. R., & Napper, L. (2005). Self-affirmation and the biased processing of threatening health-risk information. *Personality and Social Psychology Bulletin*, *31*, 1250–1263.

Hofmann, W., & Wilson, T. D. (2010). Consciousness, introspection, and the adaptive unconscious. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 197–215). New York, NY, US: Guilford Press.

Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A tutorial on testing hypotheses using the Bayes factor. *Psychological Methods*. https://doi.org/10.1037/met0000201

Huynh, S., Stefanucci, J. K., & Aspinwall, L. G. (2014). Self-affirmation counters the effects of self-regulatory resource depletion on height perception. *Journal of Experimental Social Psychology*, *52*, 96–100.

Inquisit 5. (2016). Millisecond Software.

Inzlicht, M., & Schmeichel, B. J. (2012). What Is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, *7*, 450–463.

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.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, *22*, 337–349.

Kirschbaum, C., Pirke, K.-M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test' – A tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*, 76–81.

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, *143*, 131–141.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*, 661–679.

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, *32*, 5.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*, 764–766.

Loseman, A., & van den Bos, K. (2012). A self-regulation hypothesis of coping with an unjust world: Ego-depletion and self-affirmation as underlying aspects of blaming of innocent victims. *Social Justice Research*, *25*, 1–13.

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.

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.

Lurquin, J. H., & Miyake, A. (2017). Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers in Psychology*, *8*, 568.

McQueen, A., & Klein, W. M. P. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, *5*, 289–354.

Miketta, S., & Friese, M. (2019). Debriefed but still troubled? About the (in)effectiveness of postexperimental debriefings after ego threat. *Journal of Personality and Social Psychology*, *117*, 282–309.

Nguyen, T., Carnevale, J. J., Scholer, A. A., Miele, D. B., & Fujita, K. (2019). Metamotivational knowledge of the role of high-level and low-level construal in goal-relevant task performance. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspa0000166

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.

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.

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Richter, M., Friedrich, A., & Gendolla, G. H. E. (2008). Task difficulty effects on cardiac activity. *Psychophysiology*, *45*, 869–875.

Richter, M., & Gendolla, G. H. E. (2009). The heart contracts to reward: Monetary incentives and preejection period. *Psychophysiology*, *46*, 451–457.

Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three decades of research on motivational intensity theory. In *Advances in Motivation Science* (Vol. 3, pp. 149–186). Elsevier.

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.

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, *22*, 322–339.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 261–302). Academic Press.

Storr, S. M., & Sparks, P. (2016). Does self-affirmation following ego depletion moderate restrained eaters' explicit preferences for, and implicit associations with, high-calorie foods? *Psychology & Health*, *31*, 840–856.

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.

van de Mortel, T. F. (2005). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing*, *25*, 10.

Wright, R. A. (1996). Brehm's theory of motivation as a model of effort and cardiovascular response. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 424–453). New York, NY, US: Guilford Press.

Wright, R. A., Greenberg, J., & Brehm, S. S. (2003). *Motivational analyses of social behavior: Building on Jack Brehm's contributions to psychology*. Mahwah, N.J: Lawrence Erlbaum.

Wright, R. A., Junious, T. R., Neal, C., Avello, A., Graham, C., Herrmann, L., … Walton, N. (2007). Mental fatigue influence on effort-related cardiovascular response: Difficulty effects and extension across cognitive performance domains. *Motivation and Emotion*, *31*, 219–231.

Wright, R. A., Mlynski, C., & Carbajal, I. (2019). Outsiders' thoughts on generating self-regulatory-depletion (fatigue) effects in limited-resource experiments. *Perspectives on Psychological Science*, *14*, 469–480.

Supplement

Lay Theories About Willpower (Exploratory)

The effect of self-control demands on Stroop interference (errors) was not moderated by participants' lay theories about willpower (α = 0.86; see Table 1).

Table 1

Lay theories about willpower: Moderation of the effect of self-control demands on Stroop interference (errors)

	b	*SE B*	*t*	*p*
Self-control demands	0.50	1.17	0.43	.667
Lay theories about willpower (exhaustion)	0.54	0.78	0.70	.487
Self-control demands × Lay theories	0.39	1.18	0.33	.744

Note. N = 174. R^2 = .01. Continuous predictors are mean-centered and scaled by 1 *SD*.

Trait Self-Control (Exploratory)

The effect of self-control demands on Stroop interference (errors) was not moderated by participants' trait self-control (α = 0.85; see Table 2).

Table 2

Trait self-control: Moderation of the effect of self-control demands on Stroop interference (errors)

	b	*SE B*	*t*	*p*
Self-control demands	0.50	1.15	0.43	.667
Trait self-control	-0.94	0.88	-1.07	.286
Self-control demands × Trait self-control	-0.72	1.17	-0.62	.535

Note. N = 174. R^2 = .03. Continuous predictors are mean-centered and scaled by 1 *SD*.

Stroop Interference Based on Reaction Times (Exploratory)

Participants in *the non-affirmed* conditions did not perform worse after exerting self-control than after not exerting self-control ($t_{\text{one-sided}}[75.16] = -1.12$, $p = .133$, $d = -0.25$, 95% CI [-0.68, 0.19]; for descriptive statistics, see Table 3). Self-affirmation did not counteract the detrimental effect of the ego depletion manipulation (pre-registered contrast: *high demands/non-affirmed* = +3, remaining three conditions = -1). The pre-registered contrast thus did not fit the data ($F[1, 166] = 0.17$, $p = .682$, $f = 0.03$).

Table 3

Stroop interference on reaction times (outlier cutoff 2.5 SDs from condition means; reaction time [ms] on incongruent trials minus reaction time [ms] on congruent trials)

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	194.15 (62.62) *N* = 41	220.16 (65.43) *N* = 42	*207.31 (65.00)*
	Low	180.46 (48.25) *N* = 43	195.77 (61.33) *N* = 42	*188.02 (55.31)*
	Overall	*187.14 (55.81)*	*207.96 (64.21)*	

Note. M (SD).

Systolic Blood Pressure

Outlier-cutoff 3 MADs (pre-registered, exploratory). As expected, participants showed no difference in SBP during the baseline (2 × 2 ANOVA with self-control demands × self-affirmation: all *p*s > .130). Also, as expected, participants in the high compared to the low demands conditions exerted more mental effort during the manipulation of self-control demands—difficult versus easy counting task—as indicated by higher SBP relative to the baseline ($t[160.03] = -4.78$, $p < .001$, $d = -0.73$, 95%CI [-∞, -0.47]). This speaks to the strength of our manipulation in terms of required mental effort and suggests that the SBP measurement was successful in picking up participants' increased effort in the high as compared to the low demand conditions.

During the self-affirmation manipulation, the four experimental conditions did not differ in invested mental effort (*self-control demands* × *self-affirmation* ANOVA: all *p*s > .118). Also, there were no differences between conditions as indicated by SBP reactivity during the Stroop task—neither as a function of self-control demands, nor as a function of self-affirmation or their interaction (all *p*s > .259; for descriptive statistics see Table 4).

Table 4

Systolic blood pressure with outlier cutoff 3.0 MADs

1) Systolic blood pressure (SBP) in mmHg

A) Mean SBP during baseline.

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	113.82 (10.13 *N* = 43	110.74 (7.56) *N* = 45	*112.25 (8.99)*
	Low	110.84 (8.74) *N* = 46	111.96 (10.27) *N* = 45	*111.40 (9.49)*
	Overall	*112.28 (9.50)*	*111.35 (8.99)*	

B) SBP Reactivity: Manipulation of self-control demands.

Self-control demands	High	6.54 (5.27) *N* = 86
	Low	3.13 (4.07) *N* = 88

C) SPB Reactivity: Self-affirmation manipulation.

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	6.50 (6.95) *N* = 43	6.35 (4.23) *N* = 45	*6.42 (5.69)*
	Low	3.96 (5.76) *N* = 44	6.24 (4.88) *N* = 43	*5.09 (5.44)*
	Overall	*5.21 (6.47)*	*6.29 (4.53)*	

D) SBP Reactivity: Stroop task.

		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	6.78 (5.99) *N* = 43	5.27 (5.32) *N* = 44	*6.02 (5.68)*
	Low	5.74 (7.01) *N* = 46	6.22 (4.91) *N* = 45	*5.98 (6.04)*
	Overall	*6.24 (6.52)*	*5.75 (5.11)*	

Note. M (SD).

Mental fatigue. There were no differences concerning self-reported mental fatigue between the conditions neither at the beginning of the study (as a function of self-control demands) nor after the manipulation of self-control demands (as a function of self-control demands and self-affirmation). For descriptive statistics see Table 5.

Table 5

Self-reported mental fatigue (mean of 5 items)

A) At the beginning of the study.		
Self-control demands	High	2.84 (1.26)
	Low	2.92 (1.12)
	Overall	*2.88 (1.19)*

B) After the manipulation of self-control demands.		Self-affirmation		
		No	Yes	*Overall*
Self-control demands	High	3.38 (1.29)	3.06 (1.15)	*3.22 (1.22)*
	Low	3.24 (1.21)	3.26 (1.12)	*3.25 (1.16)*
	Overall	*3.31 (1.24)*	*3.16 (1.13)*	

Note. M (SD).

PART III

Do People Avoid Mental Effort After Facing a Highly Demanding Task?

Karolin Gieseler
Saarland University

Michael Inzlicht
University of Toronto

Malte Friese
Saarland University

Manuscript submitted to the *Journal of Experimental Social Psychology*.
(Currently under revision)

Abstract

Ego depletion effects are usually examined in a sequential task paradigm in which exerting mental effort in a first task is thought to affect performance on a subsequent self-control task. A so-called ego depletion effect is observed if performance on the second task is impaired for the high demand relative to the low demand group. The present studies take a different approach. Instead of measuring performance in the second task that is equally difficult for all participants, the present studies investigated effects of self-control exertion on the self-directed selection of mental effort in the second task. Three pre-registered studies investigated if participants select less effort demanding math problems for upcoming tasks compared to a control group after exerting mental effort in an initial task. Results were mixed. Study 1 ($N = 86$) revealed no significant effect of mental effort exertion on mean choice difficulty. In Study 2 ($N = 269$), the expected effect emerged when controlling for math self-efficacy and math ability, both robustly associated with the choice measure. Study 3 ($N = 330$) descriptively replicated this result. An internal random-effects meta-analysis revealed a small overall effect of $g = 0.18$ when accounting for math self-efficacy and self-rated math ability, albeit with large heterogeneity. Exploratory analyses point to the importance of the subjective experience of mental effort in effort-selection paradigms. We discuss the potential implications of the small overall effect size for effort choice in everyday life.

Keywords: effort choice, self-control, ego depletion, demand selection

Do People Avoid Mental Effort After Facing a Highly Demanding Task?

Effects of self-control exertion on subsequent behavior are usually examined using a sequential task paradigm: Participants engage in a task that is either high or low in self-control demands. Subsequently, self-control performance is measured in a second task. If participants in the high demand group show impaired performance relative to participants in the low demand group, a so-called ego depletion effect is observed. The present studies leave this beaten track. Rather than assessing self-control performance in the second task, the present studies examined whether the exertion of mental effort influences the self-imposed choice of mental effort. Specifically, we tested the hypothesis that after high demand people would avoid mental effort by choosing less demanding variants of the second task.

Self-control is defined as "…the 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). According to the *ego depletion phenomenon*, people who exert self-control in a first task perform poorer on subsequent self-control demanding tasks.

Several hundred studies seemingly support the ego depletion idea (see Hagger, Wood, Stiff, & Chatzisarantis, 2010, for an early meta-analysis). In recent years however, research investigating ego depletion effects has been heavily criticized based on conceptual and empirical deficiencies and difficulties to replicate (e.g. Carter, Kofler, Forster, & McCullough, 2015; Carter & McCullough, 2014; Hagger et al., 2016; Osgood, 2017; Lurquin et al., 2016; Gieseler, Loschelder, & Friese, 2019). These discussions culminated in questioning the very existence of ego depletion effects (for an overview, see Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2019).

Ego depletion effects have mostly been investigated using the so-called sequential task paradigm: Participants engage in a first task requiring a little versus a lot of self-control; the second task measures participants' maximum self-control performance on tasks with fixed difficulty such as interference in a Stroop task or performance on the hand grip task. Using maximum self-control performance in tasks with fixed difficulty as the dependent variable is *one* possible— but possibly not always optimal—strategy to examine ego depletion effects. For example, ego depletion effects may be masked when participants realize that performance is key and mobilize extra effort to perform well.

Furthermore, ego depletion effects may manifest differently than in impaired maximum self-control performance. In daily life, people are often free to choose their next activity after completing a demanding task, potentially avoiding continued high demands. There may be a difference between what people are able as opposed to willing to show after exerting self-control. The implication for laboratory work is that the tendency to selectively seek out versus avoid mentally demanding activities after the exertion of self-control may be an alternative viable and potentially subtler indicator of ego depletion effects.

Theoretical and Empirical Grounds for the Hypothesis

Several theoretical models support the prediction that after an initial demanding task people will tend to prefer less demanding activities instead. The process model of self-control (Inzlicht & Schmeichel, 2012; Inzlicht, Schmeichel, & Macrae, 2014) assumes that people are less willing to exert effort after initially engaging in demanding tasks. Instead, they seek to do something more pleasurable and rewarding. This suggests that after exerting effort, people will invest less in a subsequent demanding task and consequently perform poorer. It also suggests that, when given the choice, people will choose to not be confronted with further demanding tasks. The opportunity cost model (Kurzban, Duckworth, Kable, & Myers, 2013) proposes that people estimate opportunity costs in terms of possible alternatives to a task at hand. If the opportunity costs for a given task are higher than the relative utility of the next best action, people should disengage and prioritize the alternative task. If given the choice between levels of increasing difficulty, people should choose easier levels if the opportunity costs of the harder levels are high—for instance because of high mental demand of the preceding task.

The labor/leisure tradeoff concept (Kool & Botvinick, 2014) makes a similar assumption: People strive for balance of labor (e.g., cognitive effort) and leisure (e.g., mind-wandering). Thus, participants who have worked on a highly demanding task should favor less demanding variants of following tasks. Finally, the revised version of the strength model of self-control (Baumeister & Vohs, 2016) could account for the same prediction. The model predicts that after engaging in demanding activities, people are inclined to conserve the precious limited self-control resource. One possible way would be to choose easier levels of the subsequent task if given the opportunity.

Apart from the substantive theoretical basis, some empirical research provides suggestive evidence for this idea. In one study, people low (compared to high) in trait self-control chose low (vs. high) demanding task options more often (Kool, McGuire, Wang, & Botvinick, 2013). In another study, participants who had engaged in a demanding task were more likely to rely on heuristics, a means to spare mental effort (Pohl, Erdfelder, Hilbig, Liebke, & Stahlberg, 2013). In a study examining hypothetical effects of exerting effort, participants who imagined [sic!] being depleted after an exhausting day chose less cognitively and emotionally demanding and more funny film alternatives compared to participants who imagined being energetic (Eden, Johnson, & Hartmann, 2018). Finally, one study suggests that self-reported effort avoidance may partly mediate ego depletion effects (Sjåstad & Baumeister, 2018). These studies provide scattered evidence for the hypothesis, even though they are mostly based on fairly small samples sizes and to date none of these findings has been replicated. As of yet, however, there has been no systematic examination of ego depletion effects on the subsequent selection of demanding versus less demanding tasks presented as such.

The Present Research

The present studies examined the assumption that engaging in a mentally demanding first task leads to the selection of less effortful subsequent activities. To this end—after the manipulation of mental demand—we provided participants with multiple choices to exert versus avoid mental effort by repeatedly letting them choose the difficulty of upcoming tasks.[1]

We expected self-reported experienced mental demand (subjective phenomenology) to be higher in the *high* as compared to the *low mental demand condition*. Furthermore, we expected participants to more frequently select easier alternatives in the *high* as compared to the *low mental demand* condition in a subsequent math effort task.[2] In all three studies, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in this manuscript or the supplementary material (Simmons, Nelson, & Simonsohn, 2012). In the main manuscript, we report only variables relevant for the pre-registered analyses.

[1] Initially, ego depletion effects were expected to occur specifically after the exertion of self-control (i.e., the inhibition of dominant response tendencies). Over the years, the range of manipulations used to elicit ego depletion effects broadened to behaviors that more generally can be described as mentally demanding (e.g., working memory tasks, Schmeichel, 2007; Inzlicht, Schmeichel, & Macrae, 2014). We therefore use the term "mental demand" as opposed to "self-control".
[2] Overall, we conducted 4 studies using this dependent variable. One study is not reported here because the subjective phenomenology measures that we used as a manipulation check revealed that the manipulation of mental demand was not successful, precluding a test of the hypothesis (osf.io/5b32j/).

Study 1

Method

Pre-registration and sharing. The pre-registration, all materials, data, and code can be accessed at osf.io/f78xa/.

Participants and design. We pre-registered to collect an initial sample of $N = 80$ and that we would then (a) assess our financial and human resources and (b) compute the Bayes Factor (BF) for the effect of ego depletion on effort choice to decide whether or not we would proceed data collection (Bayes Factor Design Analysis, Schönbrodt & Wagenmakers, 2018). The empirical $BF_{01} = 0.52$ indicated that the data were 1.92 times more likely under the null than under our directed hypothesis (Ly, Verhagen, & Wagenmakers, 2016). As our feasibility limit given time and resources was reached, we tested all persons who had already registered and stopped data collection. We did not run any frequentist analyses before terminating data collection.

The final sample consisted of 86 participants (98.85 % students, no psychology students, $M_{age} = 23.69$, $SD_{age} = 3.62$, 79.07 % female). We used a between participants design with two conditions (*high* vs. *low mental demand*). Sensitivity analyses using G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) revealed a statistical power of $1\text{-}\beta = .80$ (.70, .60) to detect an effect of $d = 0.54$ (0.47, 0.41) or larger (directional hypothesis).

Procedure. Up to four participants were assigned to the same experimental condition. First, they filled in an informed consent and started the experiment via the experimental software (*Inquisit 5*, 2016). They then read the general instructions, agreed to participate conscientiously and filled in several questionnaires including the first assessment of mental fatigue. After a short practice phase for the second task (dependent variable), the instructions for the first task (manipulation of mental demand) were given both orally and in writing. After this first task, we assessed mental fatigue and the measure of subjective phenomenology that also served as a check of the manipulation. This was followed by the dependent task and questions concerning mental fatigue, subjective effort and concentration during the same. Finally, participants worked through final questionnaires, including math self-efficacy and ability. Then they were payed (5€/35-40min) and thanked.

Mental demand manipulation. Participants watched funny film clips of 11 minutes (self-created compilation). Participants in the *high mental demand* condition were instructed to suppress all felt and expressed emotions while participants in the *low mental demand* condition were instructed to watch the clips like at the movies (Dang, 2018; Friese, Binder, Luechinger, Boesiger, & Rasch, 2013).[3]

Measured variables.

Math effort task. The math effort task consisted of 60 addition problems. Each problem comprised four numbers that were displayed one-by-one in the middle of the screen while participants had to update the sum in their head. Each number was presented for one second (inter-stimulus-interval [ISI] = 500ms). When the last number disappeared, participants were given a maximum of 10.3 seconds to enter the sum of the presented numbers (time specifications based on Engle-Friedman et al., 2003).

There were five levels of increasing difficulty presenting numbers in the range of 0-2 (Level 1), 2-8, 6-13, 10-25, and 12-35 (Level 5). Participants chose the difficulty level for each upcoming block of 3 problems, respectively, resulting in 20 choices (see Figure 1). Accuracy

---

[3] The video is available upon request to the first author. Due to copyright reasons we cannot make the video publicly available.

feedback was not provided. The primary dependent variable was the mean difficulty-level across the 20 choices.

At the beginning of the study, prior to the mental demand manipulation, participants worked on five practice problems, one of each level, to form an impression of the difficulties. During this practice phase, feedback was provided after each problem ("correct"/"false").

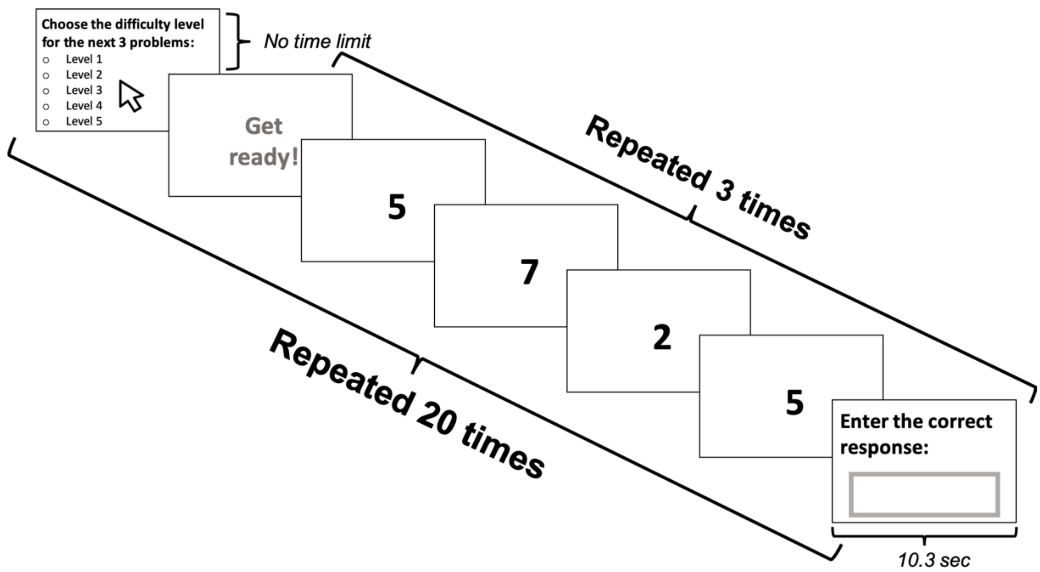


Figure 1. Flow diagram of the math effort task. Participants chose the difficulty level for each batch of three addition problems. In total, they made 20 decisions, which corresponds to 60 addition problems. The numbers shown for illustrative purposes correspond to Level 2. Numbers were presented for 1sec, ISI = 500ms.

Subjective phenomenology. After the emotion suppression task, participants answered the questions "How exhausting did it feel to work through the task?" and "How much did you have to concentrate during the task?" as compared to a 30-minute math exam (1 = *much less than in an exam* to 7 = *just like in an exam*; $\rho = 0.83$[4]; see Table 1A).

Math self-efficacy. Participants indicated their math self-efficacy on two items: "In general, I am confident that I can add several two-digit numbers in my head" and "In general, I'm good at adding several two-digit numbers in my head" (1 = *strongly disagree* to 7 = *strongly agree*; $\rho = 0.88$; see Table 1A).

Self-rated math ability. Participants rated their math ability on three items (e.g., "I am good at math", 1 = *strongly disagree* to 7 = *strongly agree*, $\alpha = .84$, see Table 1A).

Results

Pre-registered analyses. As pre-registered, we did not exclude any data points from the analyses as all participants completed the study. For pre-registered exploratory analyses, see the supplementary material.

Subjective phenomenology (confirmatory). As expected, self-reported mental effort was higher in the *high mental demand* compared to the *low mental demand* condition ($t_{one\text{-}sided}(83.29) = -3.89$, $p < .001$, $d = -0.84$, 95%CI [-1.28, -0.39]; see Table 1A).

Effort choice (confirmatory). Against our expectations, participants in the *high mental demand* condition did not select easier levels than participants in the *low mental demand* condition, but even slightly more difficult levels ($M_{\text{low mental demand}} = 2.82$, $SD_{\text{low mental demand}} = 0.61$, $M_{\text{high mental demand}} = 2.93$, $SD_{\text{high mental demand}} = 0.67$; $d = -0.16$, 95%CI [-0.59, 0.27]; see Table 1B).

[4] Spearman-Brown reliability coefficient, recommended for two item scales (Eisinga, Grotenhuis, & Pelzer, 2013).

Exploratory analyses.

Math self-efficacy. After excluding two multivariate outliers (studentized residuals > |3|, all others < |2|), there was an interaction between the mental demand condition and math self-efficacy, $b = 0.33$, 95%CI [0.12, 0.54], $t = 3.13$, $p = .002$ (see also Figure 2 & Table 1C): For people in the *low mental demand* condition, math self-efficacy was largely unrelated to effort choice ($b = 0.06$, 95%CI [-0.03, 0.16], $t = 1.34$, $p = .184$). In the *high mental demand* condition, higher math self-efficacy was associated with the selection of more difficult problems in the math effort task ($b = 0.27$, 95%CI [0.04, 0.18], $t = 6.09$, $p < .001$).

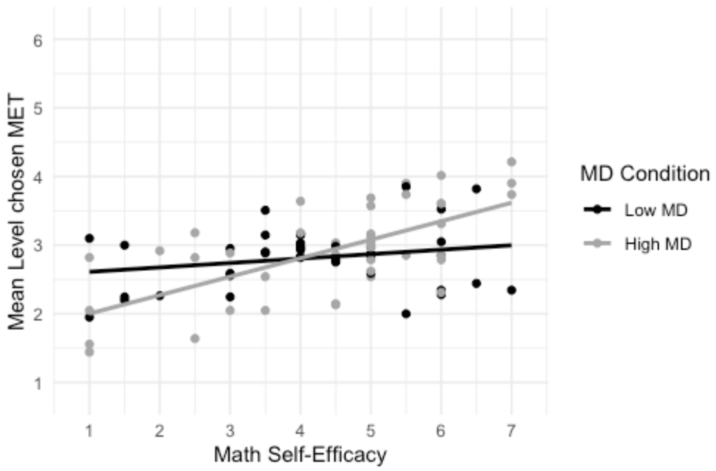


Figure 2. Self-reported math self-efficacy moderated the relationship between mental demand condition and the mean level chosen in the math effort task. Two multivariate outliers were excluded from the analysis.

Controlling for math self-efficacy and math ability. Choices in the math effort task were robustly correlated with both self-reported math self-efficacy and math ability (see Table 1D). Thus, these variables may exert an influence on the dependent variable that is not of interest to the present research question. We therefore repeated the main analysis controlling for both math self-efficacy and math ability by first residualizing math effort task choices of these scores. This analysis also revealed no significant difference in effort choices as a function of experimental condition (d = -0.06, 95%CI [-0.49, 0.37], see Table 1A).

Table 1

Study 1

A) Descriptive statistics

Condition	*N*	Subjective phenomenology (mean of 2 items)	Math self-efficacy (mean of 2 items)	Self-rated math ability (mean of 3 items)	Math effort task (mean level chosen)	Residuals Math effort task[b]
Low mental demand	42	2.19 (1.54)	4.21 (1.60)	3.77 (1.55)	2.82 (0.61)	-0.02 (0.58)
High mental demand	44	3.58 (1.77)	4.44 (1.63)	4.27 (1.64)	2.93 (0.67)	0.02 (0.51)
		$\rho = 0.83$[a]	$\rho = 0.88$	$\alpha = .84$		

Note. M (SD).
[a] Spearman-Brown reliability coefficient, recommended for two item scales (Eisinga, Grotenhuis, & Pelzer, 2013). [b] controlling for Math self-efficacy and Self-rated math ability.

B) Math effort task: Percentage of choices for different levels

Condition	*N*	Level 1	Level 2	Level 3	Level 4	Level 5
Low mental demand	42	14.01 (16.60)	25.68 (17.17)	32.67 (18.82)	19.13 (18.08)	8.51 (15.76)
High mental demand	44	14.75 (16.31)	23.81 (19.34)	25.97 (13.91)	24.66 (21.01)	10.80 (9.93)

Note. M (SD).

C) Math self-efficacy: Moderation of the effect of mental demand condition on math effort choice

	b	*SE B*	*t*	*p*
Mental demand condition	.07	.10	0.65	.516
Math self-efficacy	.10	.08	1.34	.184
Mental demand condition × math self-efficacy	.33	.11	3.13	.002

Note. $N = 84$. $R^2 = .33$. Two multivariate outliers excluded: studentized residuals > \|3\|, all others < \|2\|. Continuous predictors are mean-centered and scaled by 1 *SD*.

D) Correlations of math effort choice (math effort task) with math self-efficacy, math ability, and subjective phenomenology

Condition	*N*	Math effort task * Math self-efficacy[a]	Math effort task * Self-rated math ability	Math effort task * Subjective phenomenology
Low mental demand	42	$r = .23$ $p = .153$	$r = .34$ $p = .027$	$r = .19$ $p = .226$
High mental demand	44	$r = .66$ $p < .001$	$r = .38$ $p = .010$	$r = -.12$ $p = .445$
Total sample	86	$r = .50$ $p < .001$	$r = .37$ $p < .001$	$r = .05$ $p = .669$

Note. [a]Two multivariate outliers excluded: studentized residuals > \|3\|, all others < \|2\|.

Discussion

Even though participants rated the emotion suppression task as more mentally demanding, participants in the *high mental demand* condition did not choose easier levels in the math effort task than participants in the *low mental demand* condition. Instead, we observed a pronounced tendency to choose the three middle categories, especially the intermediate level (see Table 1B). Providing an intermediate option possibly masked differences in effort choice between the experimental conditions as people generally tend to prefer the middle options (Missbach & König, 2016; Simonson, 1989). Providing an even number of difficulty levels and thereby forcing participants to choose between options that are clearly on the more versus less effortful half of options may circumvent this issue.

Exploratory analyses revealed a moderation indicating that for participants in the *high* (but not *low*) *mental demand* condition, math self-efficacy was related to math effort choice: Those lower in math self-efficacy chose easier problems than those high in math self-efficacy. This moderation effect should be interpreted with caution. It was unexpected and the sample size of Study 1 was small.

Study 2

Study 2 addressed several issues of Study 1. First, we added a sixth category to the math effort task to circumvent the tendency to choose the middle option. Second, we sought to explore whether the unexpected finding of math self-efficacy guiding math effort choice for those in the *high,* but not *low mental demand* condition would replicate. Third, we recruited a much larger sample compared to Study 1 that allowed for adequate statistical power for smaller effect sizes.

Method

Pre-registration and sharing. The pre-registration, all materials, data, and code can be accessed at osf.io/sk9e5/.

Participants and design. We pre-registered to collect a sample of 300 participants on Amazon's Mechanical Turk. Participants received $3.60 for their participation. Although we put emphasis on our age restrictions in the study description, 31 participants indicated being older than our pre-registered criterion (18 to 35 years) and had to be excluded. The final sample thus consisted of $N = 269$ participants ($M_{\text{age}} = 29.13$, $SD_{\text{age}} = 3.86$, 46.47% female). A sensitivity analysis ($N_{\text{high mental demand}} = 144$, $N_{\text{low mental demand}} = 125$) using G*Power (Faul et al., 2007) revealed a statistical power of $1\text{-}\beta = .80$ (.70, .60) to detect an effect of $d = 0.30$ (0.27, 0.23) or larger (directional test).

Procedure. The overall procedure was the same as in Study 1. We made some modifications to address the online environment of the study: We manipulated mental demand using a different task and used slightly different items for subjective phenomenology.

Manipulated variable.

Symbol counter task. We used the symbol counter task to manipulate mental demands. The task prompts participants to count the number of small and big squares presented sequentially on the screen and requires executive control (Garavan, 2000). We used a modified version of the task to continuously adapt the difficulty for each participant based on performance (Lin, Saunders, Friese, & Inzlicht, 2019). Each trial consisted of a series of small and big squares presented sequentially on the screen. During the first trial, eleven squares were presented one-by-one separated by a fixation cross. Squares of each category (small or big) were presented in mixed order. During the first trial, there were two switches between categories (i.e., from *small* to *big* or from *big* to *small*). If participants managed to keep track of the number of squares in each category and responded correctly, the difficulty for the next trial increased. The total number of squares presented increased by one, square display time decreased by 20ms, and the number of switches increased. If participants gave an incorrect response, the total number of squares in the next trial decreased by 1, each square was presented 20ms longer and the switch frequency decreased. The task lasted eight minutes. The *low mental demand* group watched a nature video for 4:30min. Recent research has shown that tasks experienced as boring (e.g. long and easy tasks) might evoke similar subjective states as cognitively demanding tasks (Milyavskaya, Inzlicht, Johnson, & Larson, 2019). Besides, several studies suggested that increasing the duration of the manipulation for the high demand condition evokes stronger effects (Sjåstad & Baumeister, 2018; Vohs, Baumeister, & Schmeichel, 2012). We therefore refrained from matching the control task in terms of duration and task category.

Measured variables.

Subjective phenomenology. Participants rated mental demand, concentration, mental effort, frustration and mental fatigue using a slider for each item (e.g., “How mentally demanding was the task?”, from “very low demand” to “very demanding”, internally coded 1–70). The mean of all 5 items was used as the indicator of subjective phenomenology.

Math effort task. The math effort task was largely the same as used in Study 1. We added a sixth difficulty level to avoid a middle category. Additionally, we modified the range of numbers for each level to increase the difficulty of the first level and to smoothen the increase in difficulty for the following levels. Participants worked on the task for 7 minutes, so the number of choices and addition problems worked on varied between participants. Each choice was made for the upcoming 2 addition problems—instead of 3 in Study 1—to increase the overall number of choices made.

Results

Pre-registered analyses.

Subjective phenomenology (confirmatory). As expected, participants in the *high mental demand* condition indicated stronger mental effort with a large effect size, which speaks to a successful manipulation (mean of 5 items; $\alpha = .91$; $t_{one\text{-}sided}(217.89) = -14.41$, $p < .001$, $d = -1.81$, 95%CI [-∞, -1.57]; see Table 2A).

Effort choice (confirmatory). Against our predictions and parallel to Study 1, participants did not select easier problems in the *high* compared to the *low mental demand* condition ($t_{one\text{-}sided}(264.25) = 0.56$, $p = .434$, $d = 0.07$, 95%CI [-0.17, 0.31]; $M_{\text{low mental demand}} = 2.59$, $SD_{\text{low mental demand}} = 1.18$, $M_{\text{high mental demand}} = 2.51$, $SD_{\text{high mental demand}} = 1.23$, see Table 2B). However, see the additional exploratory analyses for a more nuanced analysis.

Moderation math self-efficacy (exploratory). Contrary to Study 1, math self-efficacy did not moderate the effect of the mental demand manipulation on effort choice, $b = 0.07$, 95%CI [-0.20, 0.34], $t = 0.53$, $p = .600$ (see Table 2C). Independent of the *mental demand condition*, participants reporting higher math self-efficacy chose more difficult levels in the math effort task.

Table 2

Study 2

A) Descriptive statistics

Condition	*N*	Subjective phenomenology (mean of 5 items)	Math self-efficacy (mean of 2 items)	Math ability (mean of 3 items)	Math effort task (mean level chosen)	Residuals Math effort task[b]
Low mental demand	125	21.41 (18.21)	43.89 (19.47)	34.53 (17.36)	2.59 (1.18)	0.17 (1.02)
High mental demand	144	49.43 (12.76)	50.20 (17.73)	42.48 (18.31)[a]	2.51 (1.23)	-0.15 (1.08)
		$\alpha = .91$	$\rho = 0.90$[c]	$\alpha = .76$		

Note. M (*SD*).
[a] Participants in the high mental demand condition reported higher math ability, $t(253.04) = -2.76$, $p = .006$, $d = -0.34$, 95%CI [-0.58, -0.10]. [b] controlling for Math self-efficacy and Self-rated math ability. [c] Spearman-Brown reliability coefficient, recommended for two item scales (Eisinga et al., 2013).

B) Math effort task: Percentage of choices for different levels

Condition	*N*	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6
Low mental demand	125	14.45 (13.56)	5.55 (5.45)	4.56 (5.05)	4.10 (5.62)	2.73 (4.17)	2.91 (4.54)
High mental demand	144	16.88 (14.77)	5.01 (5.19)	4.00 (5.45)	2.64 (3.77)	2.21 (3.06)	3.71 (5.57)

Note. M (SD).

C) Math self-efficacy: Moderation of the effect of mental demand condition on math effort choice in math effort task

	b	*SE B*	*t*	*p*
Mental demand condition	-.25	.14	-1.83	.069
Math self-efficacy	.47	.10	4.83	<.001
Mental demand condition × math self-efficacy	.07	.14	0.53	.600

Note. $N = 269$. $R^2 = .16$. Continuous predictors are mean-centered and scaled by 1 *SD*.

D) Correlations of math effort choice (math effort task) with math self-efficacy, math ability, and subjective phenomenology

Condition	*N*	Math effort task * Math self-efficacy	Math effort task * Math ability	Math effort task * Subjective phenomenology
Low mental demand	125	$r = 0.41$ $p < .001$	$r = 0.47$ $p < .001$	$r = -0.08$ $p = .362$
High mental demand	144	$r = 0.41$ $p < .001$	$r = 0.46$ $p < .001$	$r = -0.34$ $p < .001$

Exploratory analyses.

Controlling for math self-efficacy and math ability. Both math self-efficacy and math ability were robustly correlated with math effort task choices (*r*s > .40, *p*s < .001, see Table 2D). As in Study 1, we therefore controlled for these variables in math effort task choices before exploring the potential effect of *mental demand condition* on effort choice. This analysis revealed the expected effect: Participants in the *high mental demand* condition chose easier levels in the math effort task than those in the *low mental demand* condition ($t_{one\text{-}sided}[265.07] = 2.55$, $p = .006$, $d = 0.31$, 95%CI [0.11, +∞]).

Moderated mediation mental demand condition on math effort task via subjective phenomenology (exploratory). As reported above, the *mental demand condition* predicted subjective phenomenology after the first task. Higher ratings of mental demand were related to choosing easier math problems, but only for the *high mental demand* condition (see Table 2D). We formally tested if the effect of *mental demand condition* through subjective phenomenology ratings on math effort choice was moderated by *mental demand condition* by running a moderated mediation model (see Figure 3). This model revealed an indirect effect of the *mental demand condition* on residualized math effort choices through subjective phenomenology. This mediation was moderated by the *mental demand condition* such that higher reported effort for the symbol counter task was associated with less difficult math effort task choices for participants in the *high*, but not in the *low mental demand* condition.

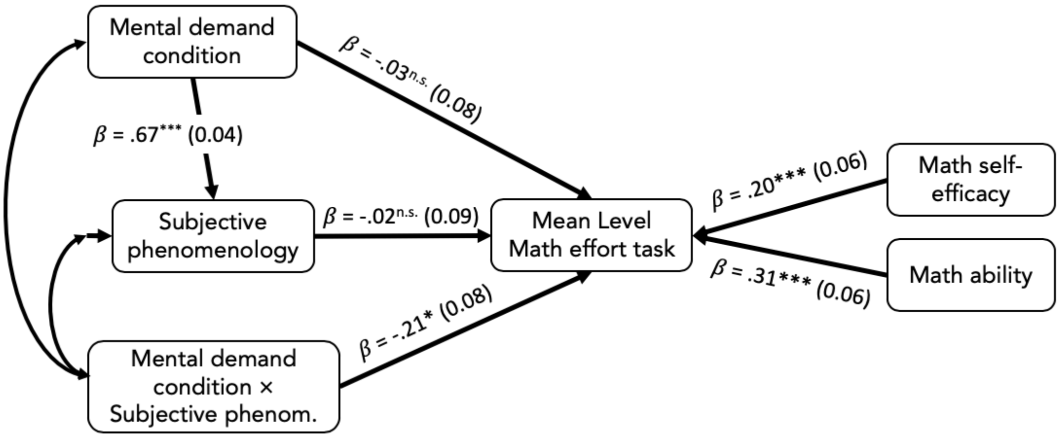


Figure 3. Moderated mediation model. β (*SE*). * $p < .05$; ***$p \leq .001$. The difference between the indirect paths for the *high* and *low mental demand* conditions was significant indicating moderated mediation (β = -0.14, *SE* = .06, p = .015).

Discussion

As in Study 1, there was no effect in the pre-registered analysis of mental demand on math effort choice. However, controlling for two constructs robustly associated with math effort choices—math self-efficacy and math ability—revealed the predicted effect: Participants in the *high mental demand* condition chose less effort-demanding levels than participants in the *low mental demand* condition. Note that this result emerged from an exploratory analysis that requires replication. The moderation effect of math self-efficacy on effort choice that we unexpectedly found in Study 1 was not replicated.

Study 3

Study 3 sought to replicate the results of Study 2 with yet a larger sample.

Method

Pre-registration and sharing. The pre-registration, all materials, data, and code can be accessed at osf.io/u9sj3/. We pre-registered Bayesian analyses in addition to frequentist analyses. Priors are based on the posterior distributions of Study 2 for the respective informed hypotheses.

Participants and design. We pre-registered to collect a sample of 350 participants on Amazon's Mechanical Turk. Twenty participants had to be excluded as they did not meet the age restrictions even though we added an entrance question to exclude participants who placed themselves in the wrong age category. Thus, the final sample consisted of 330 participants (M_{age} = 29.54, SD_{age} = 3.88, 41.52% female). They received $3.60 for their participation. A sensitivity analysis for the final sample ($N_{high\ mental\ demand}$ = 162, $N_{low\ mental\ demand}$ = 168) using G*Power (Faul et al., 2007) revealed a statistical power of 1-β = .80 (.70, .60) to detect an effect of d = 0.27 (0.24, 0.21) or larger (directional test).

Procedure. The overall procedure was the same as in Study 2. We made only minor changes: We added an item assessing boredom after the subjective phenomenology items. To further increase the variance in math effort choice, we split the first level of the math effort task into two levels. Level 1 was picked most often in Study 2. Besides, we added a question to assess self-rated data quality. For a full list of measured variables, see the supplementary material.

Results

Pre-registered analyses.

Subjective phenomenology (confirmatory). As expected, participants in the *high mental demand* condition rated the first task as more demanding than those in the *low mental demand* condition (mean of 5 items; $t_{one\text{-}sided}$ [319.30] = -20.88, p < .001, d = -2.29, 95%CI [-∞, -2.06]; BF_{10} = 9.71e^{58}, pre-registered normal prior: M = -1.77, SD = 0.26; for descriptive statistics see Table 3A).

Effort choice (confirmatory). We first residualized math effort task scores by math self-efficacy and self-rated math ability. Descriptively, participants in the *high mental demand* condition chose less effort-demanding alternatives in the math effort task compared to those in the *low mental demand* condition. However, this effect was not significant and only about half as large as in Study 2 ($t_{one\text{-}sided}$[327.97] = 1.31, p = .096, d = 0.14, 95%CI [-0.07, 0.36]; BF_{10} = 1.10, normal prior: M = 0.26, SD = 0.18; for descriptive statistics see Table 3A & B).

Moderated mediation of mental demand condition on math effort choices via subjective phenomenology (exploratory). As in Study 2, the effect of the manipulation of mental demand on math effort choice was mediated by the subjective phenomenology. However, in Study 3 the mediation was not moderated by *mental demand condition*. For both conditions, reporting higher ratings in the subjective phenomenology resulted in the choice of easier levels in the math effort task (see Table 3C, Figure 4).

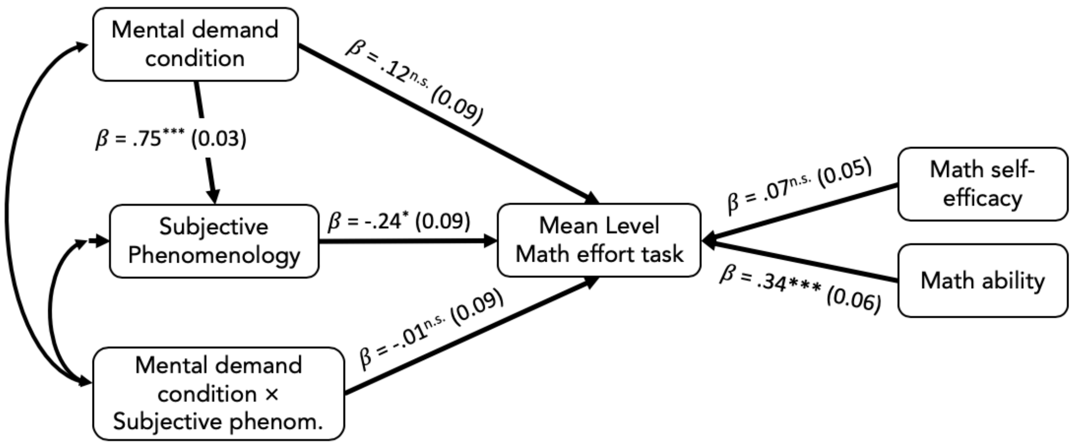


Figure 4. Mediation model. β (SE). * $p < .05$; *** $p \leq .001$. Subjective phenomenology moderates the effect of *mental demand condition* on math effort choice. Contrary to Study 2, there was no difference between the indirect path for *high* and *low mental demand* when including *mental demand condition* as additional moderator of the mediation (β = -.01, SE = .07, p = .897).

Table 3

Study 3

A) Descriptive statistics Study 3

Condition	*N*	Subjective phenomenology (mean of 5 items)	Math self-efficacy	Math ability	Math effort task (mean level chosen)	Residuals Math effort task[a]
Low mental demand	168	17.52 (15.36)	48.40 (17.03)	38.30 (16.07)	3.00 (1.56)	0.1 (1.42)
High mental demand	162	49.68 (12.53)	47.91 (18.16)	38.53 (16.05)	2.8 (1.45)	-0.1 (1.35)
		$\alpha = .93$	$\rho = 0.93$[b]	$\alpha = .74$		

Note. M (*SD*).
[a] controlling for Math self-efficacy and Self-rated math ability. [b] Spearman-Brown reliability coefficient, recommended for two item scales (Eisinga et al., 2013).

B) Math effort task: Percentage of choices for different levels

Condition	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7
Low mental demand	12.58 (14.04)	5.80 (7.78)	4.45 (6.43)	3.82 (5.49)	2.98 (4.57)	1.74 (2.86)	3.20 (6.57)
High mental demand	14.41 (15.05)	5.77 (7.38)	4.62 (6.14)	4.16 (5.81)	2.52 (3.38)	1.65 (2.95)	2.75 (6.21)

Note. M (SD).

C) Correlations of math effort choice (math effort task) with math self-efficacy, math ability, and subjective phenomenology

Condition	*N*	Math effort task * Math self-efficacy	Math effort task * Math ability	Math effort task * Subjective phenomenology
Low mental demand	168	$r = .27$ $p < .001$	$r = .42$ $p < .001$	$r = -.19$ $p = .014$
High mental demand	162	$r = .24$ $p = .002$	$r = .35$ $p < .001$	$r = -.20$ $p = .010$

Internal Meta-Analytic Summary

Multi-study papers presenting partly inconsistent findings profit from reporting internal meta-analyses. Internal meta-analyses strengthen conclusions pertinent to effect sizes, reliability and replicability of the findings—especially if effects are small—as the results are based on a larger sample than the individual studies with overall larger statistical power (Goh, Hall, & Rosenthal, 2016; Maner, 2014). We therefore summarized the findings of the three studies in an internal meta-analysis. We used random effects in which the mean effect size for the effect of mental demand on math effort choice was weighted by sample size. All effect sizes were converted to Hedge's *g* and we controlled for math self-efficacy and math ability in math effort choice in all studies. The overall effect of mental demand on math effort choice was significant with a small effect size, $z = 2.27$, $p = .023$, $g = 0.18$, 95%CI [0.02, 0.34][5] (see Figure 5). Thus, across three studies, after engaging in a highly demanding task, participants selected less effort-demanding math problems when the influence of math self-efficacy and self-rated math ability was controlled for.

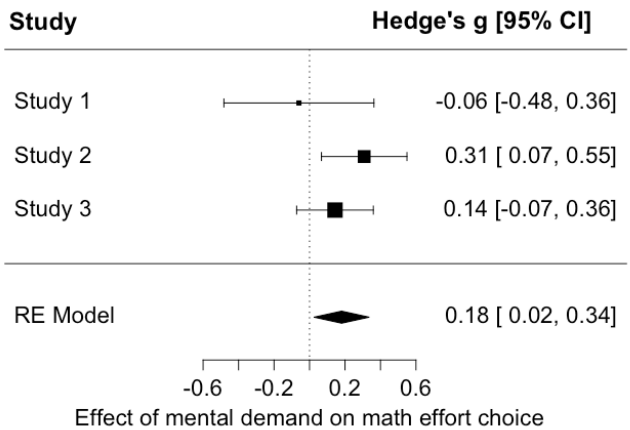


Figure 5. Internal meta-analysis of all three studies. Small average effect size of $g = 0.18$ with strong (unexplained) heterogeneity across studies.

[5] The fixed effects model yielded a similar result, $z = 2.37$, $p = .018$, $g = 0.18$, 95%CI [0.03, 0.33].

General Discussion

Ego depletion effects have usually been investigated by comparing performance on a self-control task with fixed difficulty for all participants after either facing low or high demands in an initial task. Taking a different approach, the present studies examined effects of mental demands on the subsequent choice of more or less effort-demanding upcoming tasks. We expected participants to choose less effort-demanding tasks after engaging in a demanding first task. This hypothesis was based on several theoretical accounts. After engaging in effort-demanding tasks, people may either lack the motivation to delve into yet more effortful tasks (Inzlicht & Schmeichel, 2012), prioritize easier tasks because of high opportunity costs (Kurzban et al., 2013), strive for leisure after labor (Kool & Botvinick, 2014), or be prone to conserve a limited resource (Baumeister & Vohs, 2016).

Overall, results were mixed. In Study 1, there was no significant effect of the mental demand manipulation on the choice of difficulty of upcoming math problems. This null effect may have been favored by participants' tendency to choose the middle category. Study 2 again found no difference in math effort choice as a function of *mental demand condition*, but when controlling for math self-efficacy and self-rated math ability—both robustly associated with the dependent variable—the expected effect emerged with a small to moderate effect size. Study 3 replicated this effect descriptively with a small effect size. An internal meta-analysis of all studies revealed a small average effect size of $g = 0.18$.

Studies 2 and 3 pointed to the potentially important role of subjective phenomenology: In Study 2, participants in the *high mental demand* condition and in Study 3 participants in both conditions who experienced the first task as more demanding chose easier task alternatives later on. These findings—to be interpreted with caution as there may well be alternative mediating variables (Fiedler, Harris, & Schott, 2018)—suggest that subjective experiences of mental demand, effort and fatigue may play an important role in prompting ego depletion effects and underline the importance of strong manipulations to trigger these subjective experiences (Friese et al., 2019; Wright, Mlynski, & Carbajal, 2019). One way to achieve this aim may be manipulations that vary the difficulty level adaptively, constantly confronting participants with a difficulty at the limit of their current maximum ability, as in Studies 2 and 3. Such manipulations make sure to challenge all participants similarly. A fixed difficulty for all participants throughout the task runs the risk of over- or undercharging some participants.

Strengths, Limitations, & Outlook

Several strengths lend credibility to our findings. First, we followed principles of open science and pre-registered all studies including the respective analysis plans, and provide open materials, data, and code. Second, we employed demand manipulations that were titrated to participants skill levels (Studies 2 and 3) and led to strong effects on the subjective phenomenology questions assessing demand, effort, concentration, and fatigue—a basic requirement that is not reliably met by studies in the field of ego depletion research (Friese et al., 2019; Hagger et al., 2010). Third, Studies 2 and 3 featured relatively large sample sizes. Combining all studies in an internal meta-analysis allowed us to test for even small effects with adequate statistical power.

Despite these strengths, there are obvious limitations. First, the meta-analytic effect size was small and there was strong (unexplained) heterogeneity across studies. An effect size of this magnitude ($g = 0.18$) is very difficult to study in the laboratory with adequate statistical power. A sensible next step would be to investigate effort choice in domains other than math to generalize the present findings. Should the findings generalize, we suggest to examine the same theoretical idea in people's everyday lives. Experience sampling methodology would allow to investigate if subjective experiences of high mental demand lead to the choice of less effortful activities in daily life. The investigation of such dynamics within persons would also be conducive to examining potentially small effect sizes with appropriate statistical power. This way it would be possible to compare laboratory and online findings to everyday life settings and to develop an idea about whether the respective statistical effect sizes may practically matter to people's lives.

Conclusion

We investigated whether people choose less effort-demanding task alternatives after engaging in mentally demanding tasks. Across three pre-registered studies, we found accumulated evidence that when taking math self-efficacy and self-reported math ability into account, people tend to select easier math problems when exerting mental effort beforehand. An internal meta-analysis revealed a small overall effect of $g = 0.18$, albeit with large heterogeneity between studies. Future research may turn to experience sampling methods to assess the subjective experience of effort, influences on effort choice and their relevance for daily life.

References

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

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

Dang, J. (2018). An updated meta-analysis of the ego depletion effect. *Psychological Research*, *82*, 645–651. https://doi.org/10.1007/s00426-017-0862-x

Eden, A., Johnson, B. K., & Hartmann, T. (2018). Entertainment as a creature comfort: Self-control and selection of challenging media. *Media Psychology*, *21*, 352–376. https://doi.org/10.1080/15213269.2017.1345640

Eisinga, R., Grotenhuis, M. te, & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*, 637–642. https://doi.org/10.1007/s00038-012-0416-3

Engle-Friedman, M., Riela, S., Golan, R., Ventuneac, A. M., Davis, C. M., Jefferson, A. D., & Major, D. (2003). The effect of sleep loss on next day effort. *Journal of Sleep Research*, *12*, 113–124. https://doi.org/10.1046/j.1365-2869.2003.00351.x

Faul, F., Erdfelder, E., Lang, 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. https://doi.org/10.3758/BF03193146

Fiedler, K., Harris, C., & Schott, M. (2018). Unwarranted inferences from statistical mediation tests – An analysis of articles published in 2015. *Journal of Experimental Social Psychology*, *75*, 95–102. https://doi.org/10.1016/j.jesp.2017.11.008

Friese, M., Binder, J., Luechinger, R., Boesiger, P., & Rasch, B. (2013). Suppressing emotions impairs subsequent Stroop performance and reduces prefrontal brain activation. *PLoS ONE*, *8*, e60385. https://doi.org/10.1371/journal.pone.0060385

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

Garavan, H. (2000). A parametric manipulation of central executive functioning. *Cerebral Cortex*, *10*, 585–592. https://doi.org/10.1093/cercor/10.6.585

Gieseler, K., Loschelder, D. D., & Friese, M. (2019). What makes for a good theory? How to evaluate a theory using the strength model of self-control as an example. In *Social Psychology in Action* (pp. 3–21). Springer.

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, *10*, 535–549. https://doi.org/10.1111/spc3.12267

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

Inquisit 5. (2016). Millisecond Software. Retrieved from https://millisecond.com

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, *7*, 450–463. https://doi.org/10.1177/1745691612454134

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

Kool, W., & Botvinick, M. (2014). A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, *143*, 131–141. https://doi.org/10.1037/a0031048

Kool, W., McGuire, J. T., Wang, G. J., & Botvinick, M. M. (2013). Neural and behavioral evidence for an intrinsic cost of self-control. *PLOS ONE*, *8*, e72626. https://doi.org/10.1371/journal.pone.0072626

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*, 661–679. https://doi.org/10.1017/S0140525X12003196

Lin, H., Saunders, B., Friese, M., & Inzlicht, M. (2019). *Strong effort manipulations reduce response caution: A preregistered reinvention of the ego-depletion paradigm*. Manuscript submitted for publication.

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

Ly, A., Verhagen, J., & Wagenmakers, E.-J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*, *72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004

Maner, J. K. (2014). Let's put our money where our mouth is: If authors are to change their ways, reviewers (and editors) must change with them. *Perspectives on Psychological Science*, *9*, 343–351. https://doi.org/10.1177/1745691614528215

Milyavskaya, M., Inzlicht, M., Johnson, T., & Larson, M. J. (2019). Reward sensitivity following boredom and cognitive effort: A high-powered neurophysiological investigation. *Neuropsychologia*, *123*, 159–168. https://doi.org/10.1016/j.neuropsychologia.2018.03.033

Missbach, B., & König, J. S. (2016). Middle choice preference and snack choice: The role of self-regulatory resources to nudge healthier food choice. *Food Quality and Preference*, *53*, 127–131. https://doi.org/10.1016/j.foodqual.2016.06.007

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.

Pohl, R. F., Erdfelder, E., Hilbig, B. E., Liebke, L., & Stahlberg, D. (2013). Effort reduction after self-control depletion: The role of cognitive resources in use of simple heuristics. *Journal of Cognitive Psychology*, *25*, 267–276. https://doi.org/10.1080/20445911.2012.758101

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General*, *136*, 241–255. https://doi.org/10.1037/0096-3445.136.2.241

Schönbrodt, F. D., & Wagenmakers, E.-J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, *25*, 128–142. https://doi.org/10.3758/s13423-017-1230-y

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 word solution* (SSRN Scholarly Paper No. ID 2160588). Rochester, NY: Social Science Research Network. Retrieved from https://papers.ssrn.com/abstract=2160588

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*, 158. https://doi.org/10.1086/209205

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

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

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

Wright, R. A., Mlynski, C., & Carbajal, I. (2019). Outsiders' thoughts on generating self-regulatory-depletion (fatigue) effects in limited-resource experiments. *Perspectives on Psychological Science*, *14*, 469–480. https://doi.org/10.1177/1745691618815654

Supplement

Study 1

Number of changes between difficulty levels in the math effort task (exploratory). There was no difference in the number of changes between participants in the *high* versus *low mental demand* condition ($t[83.9] = -0.81$, $p = .418$, $d = -0.18$, 95%CI [-3.36, 1.41]; $M_{\text{low mental demand}} = 13.5$, $SD_{\text{low mental demand}} = 5.54$, $M_{\text{high mental demand}} = 14.5$, $SD_{\text{high mental demand}} = 5.59$).

Performance in the math effort task (exploratory). Participants performance decreased as the difficulty level in the math effort task increased (see Table 1 & Figure 1).

Table 1

Performance math effort task (% correct)

Condition	*N*	Level 1	Level 2	Level 3	Level 4	Level 5
Low mental demand	42	98.58 (3.20)	91.55 (10.96)	78.93 (16.37)	55.88 (22.55)	37.28 (32.81)
High mental demand	44	96.75 (6.82)	90.91 (8.07)	86.30 (13.03)	62.41 (26.06)	45.35 (30.35)

Note. M (*SD*).

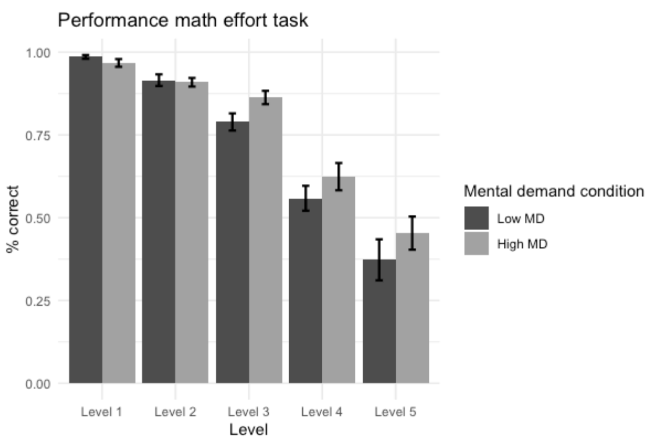


Figure 1. Performance in the math effort task as a function of difficulty level and mental demand condition. Error bars represent ± 1SEM.

Lay theories about willpower (exploratory). The effect of mental demand condition on math effort choice was not moderated by participants' lay theories about willpower (see Table 2).

Table 2

Lay theories about willpower: Moderation of the effect of mental demand condition on math effort choice

	b	*SE B*	*t*	*p*
Mental demand condition	0.11	0.14	0.78	.437
Lay theories about willpower (exhaustion)	-0.14	0.09	-1.60	.114
Mental demand condition × Lay theories	-0.05	0.14	-0.40	.692

Note. N = 86. R^2 = .08. Continuous predictors are mean-centered and scaled by 1 *SD*.

Trait self-control (exploratory). The effect of mental demand condition on math effort choice was not moderated by participants' trait self-control (see Table 3).

Table 3

Trait self-control: Moderation of the effect of mental demand condition on math effort choice

	b	*SE B*	*t*	*p*
Mental demand condition	0.11	0.14	0.82	.416
Trait self-control	-0.07	0.10	-0.65	515
Mental demand condition × Trait self-control	0.01	0.14	0.80	.936

Note. N = 86. R^2 = .02. Continuous predictors are mean-centered and scaled by 1 *SD*.

General Discussion

The present research sought to contribute to the research field of self-control depletion in three ways. First and with regard to substantiated and well-founded doubts concerning the strength and the very existence of the effect, my colleagues and I investigated the feasibility to study ego depletion effects within participants using a doubled sequential task paradigm. Within-participant-designs allow for increased statistical power without increasing the sample size. We intended to build on this initial line of research for the development of subsequent studies. Second, we turned to possible underlying mechanisms for ego depletion effects as proposed by the strength model (Baumeister & Vohs, 2016) and the process model (Inzlicht & Schmeichel, 2012). We intended to contrast the opposing predictions of the two models using the traditional sequential task paradigm in combination with a moderator approach and psychophysiological indicators of mental effort. Finally, we modified the dependent measure assessing the effects of high mental demands as opposed to low mental demands on the subsequent task. Traditionally, the dependent measure assessed maximum self-control performance using the same equally difficult task for all participants independent of prior exertion of mental demands. Based on the reflection that in daily life people are often free to choose the amount of mental effort they are willing to invest, we provided participants with multiple opportunities to choose between task of various levels of difficulty. We thus sought to answer three main questions: a) whether self-control depletion effects can be examined within participants allowing for increased statistical power without increasing the sample size, b) whether the strength or the process model approximate the underlying mechanisms of self-control depletion effects better and c) whether—if given the choice—participants choose easier tasks after facing high versus low mental demands beforehand.

Summary and Limitations of the Present Research

With regard to the first question (Part I), our findings indicate that the tentative answer is no: Ego depletion effects are not easy to examine within participants in the laboratory using a doubled sequential task paradigm. Although we consistently found the traditional ego depletion effect in the first half of each of the two studies, in the second half of each study there were no differences in self-control performance in function of *high* as opposed to *low* self-control demands. There are several limitations to this pair of studies. First, the recreational period was not as relaxing as we assumed: Participants indicated higher mental effort at the beginning of the second half of the study than at the beginning. We would have expected them to be as mentally fresh after the break between the first and the second half of the experiment as at the start. Several modifications of the break should hence be considered in future research. For instance, the recreational period could be extended. Our recreational period lasted about 8 minutes in each study. This length was based on previous research suggesting that a 10 minutes interval was long enough to cancel out detrimental effects of prior self-control exertion (Tyler & Burns, 2008). However, a large-scale study on school children and their exam results indicated that a 20- to 30-minutes break was beneficial for performance in exams (Sievertsen, Gino, & Piovesan, 2016). A longer break may thus be preferable, although it should not extend the overall time of the study too much as this would diminish the advantages of the within design in one experimental session. Besides, alternative tasks could be given to the participants during the break. Based on previous research, we chose funny and relaxing videos, gave explicit instructions to relax and provided participants with autonomy in choice to re-establish pre-study conditions (Muraven, Gagné, & Rosman, 2008; Tyler & Burns, 2008). Additional possibilities could be physical activity, providing participants with refreshing drinks or a situational reset: Giving participants the possibility to leave the room, to take a few steps and to re-enter the laboratory may facilitate the feeling of a fresh start and diminish the influence of the first half of the experiment. A second limitation of this pair of studies was that participants used strategies circumventing self-control demands of the Stroop task. Participants reported using various strategies enabling them to fulfill the task to ignore the literal meaning of the words and just to indicate the font color without reading the words themselves (e.g., by focusing on the bottom of the screen). We addressed this shortcoming in Study 2 by introducing a new rule that would make participants read the words. We cannot be sure, however, if it forced participants to deal with the intended high self-control demands or if they found other strategies to circumvent them. Indeed, a third limitation to both studies were strong learning effects: Participants performance in the dependent task improved from the first to the second half of the experiment. These learning effects possibly masked differences between the two experimental conditions that may have emerged through the manipulation of self-control demands.

The findings of the first part of this thesis can be summed up as follows: To increase statistical power, researchers in ego depletion research should collect samples of appropriate size based on a priori power calculations and conservative estimates of the underlying true effect or the smallest effect size of interest (Lakens, 2014). If intending to use within-participants design in the laboratory in one experimental session, researchers should think of a recreational period that is appropriate in length and content. Besides, they should make sure they use dependent measures that are not influenced by learning effects, but are nonetheless comparable.

With regard to the second question whether the strength or the process model receive more empirical support when combining the sequential task paradigm with a moderator and psychophysiological indicatory of mental effort, we faced an unexpected situation. As the moderator did not counteract the (mildly) detrimental effect of high self-control demands, we were not able to disentangle the predictions of the two models as planned. Moreover, the results of one of our psychophysiological measures, the pre-ejection period, raised doubts as to its validity. Nevertheless, our data indicated that we successfully manipulated self-control demands—both as indicated by a questionnaire measure and systolic blood pressure. Mental effort invested during the dependent task (as indicated by systolic blood pressure) was more in line with the predictions of the strength model. At the same time, one possible explanation of the unexpected effect of the moderator self-affirmation was more consistent with the predictions of the process model. The results pertaining to possible underlying mechanisms evoking ego depletion effects thus remain inconclusive.

With regard to the third question whether—if given the choice—participants select easier tasks after facing high versus low mental demands beforehand (Part III) our findings were heterogeneous. When no third variables were considered, there were no differences in the choice of mental effort as a function of *high* versus *low* mental demands in the preceding task. A small average effect of $g = 0.18$, however, suggested that people tend to choose less demanding math task alternatives when taking their self-efficacy and self-rated math ability into account. Besides, in two of three studies people who reported that they experienced the first task as more effortful chose easier task alternatives when given the choice. Subjective experience thus seems relevant to effort choice. Limitations to this set of studies arise from the focus on the math domain as findings may not generalize to other domains. The overall meta-analytical effect size was small and the individual effect sizes were heterogeneous. Also, effort avoidance may not always be explicit: People may feel drawn to less or more demanding task alternatives out of a gut feeling without being aware of it or able to name reasons for their choice.

An additional, more fundamental notion that concerns all of the studies presented is the question of "group-to-individual generalizability" (Fisher, Medaglia, & Jeronimus, 2018). Psychological research today mostly relies on at best large samples and dependent measures are aggregated to statistically test the hypotheses in question. The results and conclusions based on this aggregated data are then applied to intraindividual processes and behavior. This inferential proceeding is grounded on the assumption that processes on the large-scale, interindividual population level generalize to the intraindividual level. However, this generalization is only permissible under specific circumstances referred to as *ergodic*. Ergodicity implies that the effects in question are equivalent across individuals (i.e., homogenous) and stable over time (i.e., stationary). For instance, for the studies included in this thesis using the Stroop task as dependent measure, ergodicity would imply that the groups' central tendencies in Stroop performance after a *high* versus *low* demanding task should approximate each individual's central tendencies in Stroop performance after a *high* versus *low* demanding task. The same should apply to the variance in Stroop performance on the group level as compared to the individual level. Research empirically investigating the plausibility of ergodicity recently concluded that the necessary assumptions are hardly ever met—in other words, "the temptation to use aggregate estimates to draw inferences at the basic unit of social and psychological organization—the person—is far less accurate or valid than it may appear in the literature" (Fisher et al., 2018, p. 8). In light of this reasoning the paradigm and the results of Part I can be looked at from a different perspective. On the one hand it might inspire future research as an attempt to test the generalizability of a phenomenon usually examined on the interindividual level to the intraindividual level. On the other hand it could be interpreted as very preliminary evidence that the most often (implicitly) assumed equivalence of inter- and intraindividual processes may not be appropriate for ego depletion effects. For all of the remaining studies reported in this thesis, my reasoning is that the evidence we provide lies somewhere on the continuum between equivalence and non-equivalence between the population and the individual level (Adolf & Fried, 2019)—but we have no way of knowing where. Future research should at least bear the limited group-to-individual generalizability in mind. Ideally, equivalence should either be tested or alternative sophisticated methods as intensive repeated measures data should be employed to estimate equivalence and to make sound statements about the main point of interest: the individual. After all, psychology is the science of behavior and "its ultimate goal is to describe and explain what people do" (Baumeister, Vohs, & Funder, 2007, p. 396).

Reflections on Psychological Research and Common Sense

As researchers in the domain of ego depletion phenomena, my colleagues and I share the experience that laypeople intuitively comprehend descriptions of ego depletion effects in daily life. They often share anecdotes from their own experience and provide vivid examples when told about our research field. This suggests that at least in laypersons' individual, subjective experience, there is something to the phenomenon. With reference to this subjective facet of ego depletion phenomena, Job and colleagues came up with the idea of implicit theories of willpower: Depending on the personal belief whether willpower is limited (or not) will people show impaired self-control performance after an experience high in self-control demands (or not; Job et al., 2010). Thus, should future research on ego depletion phenomena focus more strongly on laypersons' subjective experiences and convictions? And in which ways should this common knowledge enter the scientific arena? Reflections regarding this and similar questions are often referred to as *commons sense*, *naïve* or *folk psychology* and discussed as to their significance for scientific psychology (Cacioppo, 2004; Heider, 1958; Malle, 2008). A popular proponent of common sense as a valuable source for scientific psychology was Heider (1958). He was convinced that “the ordinary person has a great and profound understanding of himself and of other people which, though unformulated or only vaguely conceived, enables him to interact with others in more or less adaptive ways” (Heider, 1958, p. 2). According to Heider, scientific psychology is unique as a science because laypersons’ intuitive knowledge of psychological phenomena and processes allows them to behave more or less adaptively in many situations. The same is not true for a layperson’s intuitive understanding of scientific physics; common sense psychology allows to behave appropriately in everyday situations as job interviews and funerals, but common sense physics do not allow to construct cars, smartphones or even atomic bombs (Heider, 1958). Psychology—and social psychology in particular—is therefore often confronted with the reproach of “obviousness” of the research findings. People outside psychology are prone to feel that the findings provided by (social-) psychological research are self-evident and already part of common knowledge. Allegations of this kind may have contributed to researchers studying and journals publishing predominantly “sexy” findings—possibly contributing to the so called “replication crisis” (Fiedler, 2017; Ioannidis, 2005). And such concerns are not new (Ring, 1967). The supposed obviousness, however, can often be explained by hindsight bias (e.g. Roese & Vohs, 2012) and by mutually contradictory theories and misconceptions held in common sense psychology. Prominent examples are “birds of a feather flock together” versus “opposites attract” (Kelley, 1992). Research on the prevalence of misconceptions in laypeoples’ knowledge of psychology suggests that myths are widespread, persisting and difficult to counter even by education (Furnham, Callahan, & Rawles, 2003; Furnham & Hughes, 2014). Nevertheless, common sense psychology can be a useful resource for scientific psychology if used diligently. This should be especially true for behavior people regularly observe in daily life in their personal surroundings, for relations with such a low degree of complexity—e.g. in time, space—that a layperson can overlook them (Kelley, 1992). As common sense psychology is considered to "include common people's ideas about their own and other persons' behaviors and about the antecedents and consequences of that behavior" (Kelley, 1992, p. 4), it is useful to discover concepts and mechanisms laypersons use to describe, explain and predict behavior relevant to their personal experience as these convictions strongly influence peoples' daily lives. In other words, common sense feeds itself from the knowledge lay people gain from their everyday experiences—and what really interests (social) psychological researchers is human behavior with its real world consequences and boundary conditions (Doliński, 2018; Fiedler, 2018; Kruglanski, Factor, & Jaśko, 2018). Using common sense as a basis for scientific psychology, however, is not without its dangers. The distinction between the research on lay beliefs and social behavioral consequences of laypersons' convictions on the one hand and the use of common-sense psychology as a resource for building scientific theories on the other is crucial and often ignored. The former takes as a premise that "lay beliefs and psychological theories exert a substantial causal influence on social judgments and behavior" (Fletcher, 1995, p. 221). This use is always justified and important for scientific psychology and especially social psychology—even if the common sense believes studied are implausible or indeed fundamentally wrong, they will nevertheless influence behavior as long as someone beliefs in them. To this large but disordered knowledge of common sense psychology, scientific psychology can add systematic analysis and non-confirmative thinking—two important aspects common sense psychology lacks as laypeople observe behavior in an unsystematic way, often using confirmatory hypothesis testing (Cacioppo, 2004).

As lay beliefs and conceptions cover such a wide area and exert such a great influence in social psychology, they generally lend themselves to the construction of theories. This second use, however, may be problematic. First, all researchers build intuitions over many years before starting life as a scientist; in other words, we all begin our scientific career packed with our own common-sense psychology accumulated through unsystematic observations and influenced by numerous biases during our lives until that day (Cacioppo, 2004). To become aware of these influences and to eliminate them as extensively as possible is a life's task for most psychologists and this issue should be addressed in the curricula at universities. Secondly and possibly as a result of insufficient awareness, common-sense psychology is often used in theory construction without questioning its validity and plausibility or even its usefulness for scientific psychology (Fletcher, 1995). In sum, the conclusion usually drawn is that the use of common-sense psychology in social psychology is curse and blessing—a blessing because lay theories are a rich resource as they guide human behavior; a curse because they also guide our reasoning and acting as social psychologists, often in a subconscious way. What does that mean for research on self-control depletion? I would argue that these reflections underline the importance of research on all kinds of lay theories and common-sense psychology in their own right. As Malle (2008, p. 164) put it: "Social psychology must chart out people's subjective perceptions and naïve-psychological assumptions, because—whether right or wrong—they are critical factors in guiding social interaction". Research on "all in your head beliefs" is scarce and may be underrepresented in self-control depletion research (Job et al., 2010; Job & Walton, 2017; Martijn, Tenbült, Merckelbach, Dreezens, & de Vries, 2002). A recent review discussing the research program on ego depletion effects that focuses on the strength model and relevant research from other researchers dedicates less than a page to this topic (Baumeister & Vohs, 2016, p. 107). The authors perceive this line of research as a “major challenge” (Baumeister & Vohs, 2016, p. 107) to the strength model, even though the research on lay theories and their influence on self-control depletion refer to research of common sense in its own right. Research on the underlying mechanisms of ego depletion effects from a scientific perspective and the development of theories and models should try to break free from common-sense psychology and therefore perceive research on lay theories not as a challenge but as a different perspective and enrichment. As described in Part III of this thesis, subjective experience can have a significant influence on effort selection as well and should be examined not only as a manipulation check but as its own mechanism (see Wright et al., 2019). Apart from investigating subjective experience, often referred to as manipulation check in self-control depletion research (see Friese et al., 2019), researchers could profit from common-sense psychology in qualitative analyses. Especially with regard to the severe critic this research field is facing, research on ego depletion phenomena may profit from taking a step back. Large-scale structured interviews may be a good starting point to learn if, when and how laypeople experience depletion effects and how they explain them. There are currently attempts to use the “wisdom of the crowds” of research experts to estimate the replicability of social scientific research claims (“the repliCATS project,” 2019). Maybe a similar project could use the wisdom of laypeople as experts of their daily lives. The results would probably be very different for different cultures (see Savani & Job, 2017, for differences between Americans and Indians concerning theories of willpower). An additional aspect that may gain more attention when building on common-sense psychology is *successful* self-control. Ego depletion research has predominantly focused on problematic behavior and self-control failure (however, see Gillebaart & de Ridder, 2015, for a different perspective). Laypeople may contribute their common knowledge on when and how they successfully exert self-control. If it was self-control failure that governed their lives, the world would be a different place. Nevertheless, people may have an easier time remembering self-control failures than successes as failures may require people to take action while successes may go unnoticed.

Turning to research not on common-sense but on scientific theories, established models on underlying mechanisms could be examined to estimate how strongly the model was influenced by the researchers' own common sense and lay theories. This might be especially difficult as researchers may have problems to control for their own experiences in the theories they built. One possible solution may be provided by so called "Big Data" approaches. As computers and algorithms do not underlie the influence of common-sense convictions, they may find interesting patterns in large data sets that lend themselves to theory development. Eventually, both research on common sense and theories inspired by researchers' common sense will have to stand up against the scientific process of testing, refining and testing again. If this is done properly without making use of researchers' degrees of freedom, wrong theories will ultimately be falsified—even if this process is especially difficult and lengthy in psychology due to the probabilistic nature of most theories (Dienes, 2008).

Future Directions

Some five or six years ago (viewed from 2019), this final section of my thesis addressing open questions and future research would have taken ego depletion phenomena for granted and based empirical and theoretical reflections on the existence of these behavioral effects. At this time, I feel that more fundamental questions need to be addressed (for reviews, see Friese et al., 2019; Milyavskaya, Berkman, & De Ridder, 2019), or we may see the predicted "disappearance of the [ego depletion] effect by 2020" (Vadillo, in press, p. 18).

First of all, more work needs to be done regarding constructs and definitions. The term "self-control" is used so broadly, that the validity of the construct is endangered—the increasing use of the term "umbrella construct" illustrates the problem of a broad range of notions gathering under the same roof (Milyavskaya et al., 2019; Moffitt et al., 2011). Ego depletion research would profit from clarifying relevant constructs and from collaborating with neighboring fields investigating fatigue (Wright et al., 2019), effort (Massin, 2017), and motivation (Richter et al., 2016), to name a few.

Laboratory research. Turning to laboratory research, this "umbrella" problem is also reflected in the wide variety of tasks used to manipulate and measure self-control in the laboratory. Do tasks as diverse as resisting delicious cookies, ignoring the literal meaning of a word and indicating the font color instead, and crossing out letters based on specific rules all lend themselves to manipulate and measure the same construct? And how long do these tasks have to last to produce effects of interest? We need to establish a consensus regarding these aspects if we want to continue to study ego depletion phenomena in the laboratory.

Many researchers would say that self-control includes inhibition and that it is effortful (see Gillebaart, 2018). But what about effortless self-control? If people manage to avoid the situation altogether in the first place, they may not need to invest any effort or to inhibit automatic behavioral tendencies. This facet is not addressed in laboratory studies as people are not given the opportunity to avoid self-control demanding tasks. Researchers may think of experimental designs that include this additional option (De Ridder, Kroese, & Gillebaart, 2018). The example of taking inhibition and effort for granted when designing ego depletion studies shows that researchers often make assumptions that have never been tested. This could reflect an instance of common sense sneaking into scientific psychology: Experiencing as effortful what they deem to be self-control, researchers may assume this characteristic without ever investigating if it holds true over situations, time and people (Milyavskaya et al., 2019).

Thus, the context is another important aspect to consider. Large inter- and intraindividual differences in the emergence of ego depletion effects in the laboratory point to the complexity of the relationships and circumstances. The strength model captivates by its parsimony (Gieseler et al., 2019)—however, “it is useful to also consider the possibility that the multivariate mechanisms underlying social behavior are almost certainly more complex than any hypothesis or model might suppose, and then simplify from this starting point” (Cacioppo, 2004, p. 117). One aspect that adds to this complexity is subjective experience. As suggested in Part III of this thesis, the perceived amount of invested effort, the subjective difficulty of the task at hand, and subjective experiences in general play an important role for self-control depletion effects (see Clarkson, Hirt, Jia, & Alexander, 2010; Clarkson, Otto, Hassey, & Hirt, 2016; Milyavskaya & Inzlicht, 2017). One possibility to address subjective experience of task difficulty is to construct the first task adaptively (see Part III). Using several different adaptive approaches for the first task manipulating self-control demands could allow to disentangle possible mechanisms eliciting ego depletion effects. For instance, frustration and invested mental effort during the first task may be possible mechanisms underlying ego depletion effects. If frustration was the underlying mechanism, adaptively adjusting the difficulty of the first task to always exceed the abilities of each participant should maximize ego depletion effects. Constantly confronting participants with a task they are not apt to should maximize participants’ frustration and lead to disengagement—and thus worse self-control performance in the second task. Similarly, if invested mental effort as indicated by systolic blood pressure or pre-ejection period (see Part II) was the underlying mechanism, adaptively designing the first task to meet participants’ abilities while keeping them at maximum performance—without overwhelming them—should maximize ego depletion effects. Participants should invest as much mental effort as individually possible and thereby the amount of mental effort still to invest in the second task would be reduced. The studies outlined underline the importance of theory and clearly formulated, testable predictions (see Friese et al., 2019).

The last aspect to mention regarding deficits of laboratory research concerns normative assumptions researchers make to define self-control failures. Researchers using the dual task paradigm assume that participants want to perform well—another common-sense assumption in need of testing—but laboratory tasks often lack personal significance, for example regarding long-term goals. It would thus be crucial for researchers to guarantee participants' engagement in the study (Milyavskaya et al., 2019). Also, researchers may question the normative assumptions that allow them to classify behavior as a self-control failure. If a participant's superordinate goal is to relax as much as possible during a study as a long day lies ahead of her/him, maximal self-control performance may not be what s/he invests.

Field research. Ego depletion research originates from the observation of everyday behavior. However, the overwhelming majority of studies addressing ego depletion effects was conducted in the laboratory. In the following, I outline several ideas on how to investigate ego depletion effects in daily life and advantages this may have.

Field research can potentially address several caveats of laboratory research mentioned in the last section. Self-control conflicts people experience in daily life are probably more personally relevant than any laboratory task, researchers do not have to invent more or less artificial paradigms to manipulate and measure self-control demands and performance, and the transfer to interventions addressing problematic behavior is not far. One possibility to link the large existing literature based on laboratory studies to daily life is to measure performance in tasks as the Stroop task in an initial lab session and to assess the behavior of interest assumed to relate to the laboratory measure in daily life using daily diary studies or ambulatory assessment. For instance, previous research suggests that Stroop performance (i.e. inhibition) is related to self-control demanding behavior in or at least relevant to real life such as alcohol consumption (Houben & Wiers, 2009), dietary resistance (Hofmann, Adriaanse, Vohs, & Baumeister, 2014), socially appropriate behavior (von Hippel & Gonsalkorale, 2005), and self-control recruitment in response to stressful events (Klein, Liu, Diehl, & Robinson, 2017). As stated before, there is a much larger variety of tasks apart from the Stroop task used in ego depletion research in the laboratory—researchers could thus investigate whether similar relations emerge for other laboratory tasks. Another approach linking laboratory and field research could be to manipulate self-control demands using the traditional tasks in the field environment in combination with relevant everyday behavior as dependent variable. As previous research investigating ego depletion effects in daily life was correlational throughout, this could provide further insights. For instance, participants with a self-set dietary goal could be randomly assigned to work on a Stroop task high or low in self-control demands before lunch. After lunch, they would fill in questionnaires assessing consumption behavior. Over the course of several days to weeks, different tasks could be used to prevent learning effects and intra- and interindividual effects could be investigated. However, this procedure may feel artificial for the participants and therefore have undesired effects to the dependent measures (i.e., hypothesis guessing). Besides, the independent task manipulating self-control demands would be as personally irrelevant in the field as in the laboratory.

Thus, it may be more promising to investigate effects of high or low self-control demands that naturally occur in people's lives, are highly self-relevant and consider individual goals. Previous correlational research closely testing the ego depletion idea suggested that higher demands in daily self-control were related to less successfully resisting desires (Hofmann, Vohs, & Baumeister, 2012), exceeding personal drinking limits (Muraven, Collins, Shiffman, & Paty, 2005), increased sweets consumption (Sonnentag, Pundt, & Venz, 2017), as well as more interpersonal conflict and neglecting responsibilities (Simons, Wills, Emery, & Spelman, 2016). Research in the context of organizational behavior provided evidence for the negative relations between smartphone use at night on work engagement (Lanaj, Johnson, & Barnes, 2014) and between helping others at work in terms and subjective feelings of depletion (Lanaj, Johnson, & Wang, 2016). Daily diary and experience sampling studies suggested that ego depletion experienced at work positively relates to spousal conflict and negatively to providing spousal support (Germeys & De Gieter, 2018) and that job stressors positively relate to subjective feelings of depletion.

How could these examples of correlational research in daily life be added a component directly manipulating self-control demands to infer causal relations? Thinking of people who are assigned certain shifts or tasks, it might be possible to (quasi-) experimentally assign them to more or less self-control demanding ones. The amount of self-control demanded by each instance could be assessed in advance and dependent measures could be related to personal goals as doing sports in the evening, food and alcohol consumption and relationship goals. For some of these goals it may even be possible to directly assess them via smartphone sensing methods without the detour via self-reports. Sleeping behavior, physical activity as running or cycling and even the content and mood of evening conversations with partners and kids could be assessed via smartphone. The same applies to behaviors relevant for household responsibilities as vacuuming and taking out the trash, to mobility patterns indicating the amount of time spend in fast food restaurants versus salad bars, and to smartphone use in general—which many people would like to reduce nowadays (for an overview, see Harari, Müller, Aung, & Rentfrow, 2017). Admittedly, manipulating self-control demands in real life may pose a major challenge. Field and laboratory research on ego depletion phenomena may therefore work best in combination profiting from one another.

As to the importance of subjective experiences in ego depletion research, combining measures of self-control performance with objective indicators of mental effort (see Part II) may be a promising option for research outside of the lab as well. Recent technologies allow to more or less continuously and non-invasively measure invested mental effort over the course of a day, for instance systolic blood pressure using a new generation of smartwatches ("HeartGuide | Wearable Blood Pressure Monitor | Omron," 2019) or heart rate using a smartphone camera (Lakens, 2013). A recent study, for instance, investigated fatigue in nurses over the course of their 12-hours shift (Johnston et al., 2019). The researchers used a combination of continuously measured physical energy expenditure (heart rate combined with activity data) with self-reports of fatigue and work demands. The results were more in line with a motivational as opposed to a resource-based account. Exerted effort was not consistently related to fatigue or perceived demands, partly due to substantial interindividual heterogeneity—experienced reward and control however predicted reduced fatigue. As in the laboratory, using this combination of performance, subjective experience and psychophysiological data has the potential to promote theorizing concerning ego depletion phenomena.

In sum, in the laboratory as in the field, a lot of empirical and theoretical work needs to be done to clarify the unsatisfying state of inconclusiveness that characterizes the current status of ego depletion research. The combination of consensus concerning constructs of interest, various kinds of variables agreed upon to manipulate and measure self-control and possibly recent technological developments may make it possible to conclude what can be said about the ego depletion idea. Open science principles and adequate statistical power are considered fundamental prerequisites.

Concluding Remarks

Self-control is considered a fundamental capacity in life, enabling us to pursue or short- and long-term goals, to resist temptations and to override automatic behavior, emotions and thoughts if we deem a different demeanor as more appropriate and desirable. Over the last two-and-a-half decades, research conducted predominantly in the laboratory using the so-called sequential task paradigm investigated the idea that people are more prone to fail at self-control demanding tasks after having faced high demands beforehand (ego depletion). Prominent models as the strength model (Baumeister & Vohs, 2016) and the process model (Inzlicht & Schmeichel, 2012) suggested a domain-general, limited resource or motivational processes as explanations for the underlying mechanism. In recent years, doubts arouse as to the very existence of the phenomenon and the research field faces severe criticism as to its conclusions, methods and theorizing. The present research tried to address a few of these issues in order to contribute to the development of the field: low statistical power and within-participant designs as a possible remedy, insufficient theorizing and possibilities to disentangle the mechanisms proposed by the two most prominent models, and alternatives to traditional dependent measures more relevant to daily life. Overall the evidence provided tentatively speaks more to the ego depletion idea than against it. However, there is a lot more empirical research—both in the laboratory and in the field—and a lot of theorizing to be done to conclude if there is something to the ego depletion idea. The studies presented underline that ego depletion effects are not easy to investigate within participants and that psychophysiological indicators of mental effort are important to shed light on the underlying mechanisms. We also found preliminary evidence speaking to the importance of subjective experiences of mental effort and to effort choice as a possibly useful alternative to the traditional self-control performance measure. Due to the potential importance of self-control depletion phenomena for our society and given the intuitive support laypeople provide for the ego depletion idea, future research should try to reach a conclusion regarding the very existence of ego depletion effects.

References

Adolf, J. K., & Fried, E. I. (2019). Ergodicity is sufficient but not necessary for group-to-individual generalizability. *Proceedings of the National Academy of Sciences*, *116*, 6540–6541. https://doi.org/10.1073/pnas.1818675116

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

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. http://dx.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., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail a self-regulation*. San Diego, CA, US: Academic Press.

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

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*, 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

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

Cacioppo, J. T. (2004). Common sense, intuition, and theory in personality and social psychology. *Personality and Social Psychology Review*, *8*, 114–122. https://doi.org/10.1207/s15327957pspr0802_4

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

Clarkson, J. J., Hirt, E. R., Jia, L., & Alexander, M. B. (2010). When perception is more than reality: The effects of perceived versus actual resource depletion on self-regulatory behavior. *Journal of Personality and Social Psychology*, *98*, 29–46. https://doi.org/10.1037/a0017539

Clarkson, J. J., Otto, A. S., Hassey, R., & Hirt, E. R. (2016). Perceived mental fatigue and self-control. In *Self-Regulation and Ego Control* (pp. 185–202). Elsevier. https://doi.org/10.1016/B978-0-12-801850-7.00010-X

Converse, P. D., & DeShon, R. P. (2009). A tale of two tasks: Reversing the self-regulatory resource depletion effect. *Journal of Applied Psychology*, *94*, 1318–1324. https://doi.org/10.1037/a0014604

Dang, J. (2018). An updated meta-analysis of the ego depletion effect. *Psychological Research*, *82*, 645–651. https://doi.org/10.1007/s00426-017-0862-x

Dang, J., Björklund, F., & Bäckström, M. (2017). Self-control depletion impairs goal maintenance: A meta-analysis. *Scandinavian Journal of Psychology*, *58*, 284–293. https://doi.org/10.1111/sjop.12371

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 Ridder, D. T. D., Kroese, F., & Gillebaart, M. (2018). Whatever happened to self-control? A proposal for integrating notions from trait self-control studies into state self-control research. *Motivation Science*, *4*, 39–49. https://doi.org/10.1037/mot0000062

Dienes, Z. (2008). *Understanding psychology as a science: An introduction to scientific and statistical inference*. Macmillan International Higher Education.

Doliński, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin*, *13(2)*, e25025. https://doi.org/10.5964/spb.v13i2.25025

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

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*, 19–23. https://doi.org/10.1111/1467-8721.00160

Engle, R. W., Conway, A. R. A., Tuholski, S. W., & Shisler, R. J. (1995). A resource account of inhibition. *Psychological Science*, *6*, 122–125. https://doi.org/10.1111/j.1467-9280.1995.tb00318.x

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

Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science*, *12*, 46–61. https://doi.org/10.1177/1745691616654458

Fiedler, K. (2018). How to make psychology a genuine science of behavior: Comment on Dolinski's thoughtful paper. *Social Psychological Bulletin*, *13(2)*, e26079. https://doi.org/10.5964/spb.v13i2.26079

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*, E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fletcher, G. J. O. (1995). Two uses of folk psychology: Implications for psychological science. *Philosophical Psychology*, *8*, 221–238. https://doi.org/10.1080/09515089508573155

Francis, Z., Milyavskaya, M., Lin, H., & Inzlicht, M. (2018). Development of a within-subject, repeated-measures ego-depletion paradigm. *Social Psychology*, 1–16. https://doi.org/10.1027/1864-9335/a000348

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., 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

Furnham, A., Callahan, I., & Rawles, R. (2003). Adults' knowledge of general psychology. *European Psychologist*, *8*, 101–116. https://doi.org/10.1027//1016-9040.8.2.101

Furnham, A., & Hughes, D. J. (2014). Myths and misconceptions in popular psychology: Comparing psychology students and the general public. *Teaching of Psychology*, *41*, 256–261. https://doi.org/10.1177/0098628314537984

Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. M., … Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, *92*, 325–336. https://doi.org/10.1037/0022-3514.92.2.325

Garrison, K. E., Finley, A. J., & Schmeichel, B. J. (2019). Ego depletion reduces attention control: Evidence from two high-powered preregistered experiments. *Personality and Social Psychology Bulletin*, *45*, 728–739. https://doi.org/10.1177/0146167218796473

Germeys, L., & De Gieter, S. (2018). A diary study on the role of psychological detachment in the spillover of self-control demands to employees' ego depletion and the crossover to their partner. *European Journal of Work and Organizational Psychology*, *27*, 140–152. https://doi.org/10.1080/1359432X.2017.1417259

Gieseler, K., Loschelder, D. D., & Friese, M. (2019). What makes for a good theory? How to evaluate a theory using the strength model of self-control as an example. In K. Sassenberg & M. L. W. Vliek (Eds.), *Social psychology in action: Evidence-based* *interventions from theory to practice* (pp. 3–21). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-13788-5_1

Gillebaart, M. (2018). The 'operational' definition of self-control. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.01231

Gillebaart, M., & de Ridder, D. T. D. (2015). Effortless self-control: A novel perspective on response conflict strategies in trait self-control. *Social and Personality Psychology Compass*, *9*, 88–99. https://doi.org/10.1111/spc3.12160

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

Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, *18*, 83–90. https://doi.org/10.1016/j.cobeha.2017.07.018

HeartGuide | Wearable Blood Pressure Monitor | Omron. (2019, June 8). Retrieved from https://omronhealthcare.com/products/heartguide-wearable-blood-pressure-monitor-bp8000m/

Heider, F. (1958). *The psychology of interpersonal relations*. Hoboken, NJ, US: John Wiley & Sons Inc. https://doi.org/10.1037/10628-000

Hofmann, W., Adriaanse, M., Vohs, K. D., & Baumeister, R. F. (2014). Dieting and the self-control of eating in everyday environments: An experience sampling study. *British Journal of Health Psychology*, *19*, 523–539. https://doi.org/10.1111/bjhp.12053

Hofmann, W., Vohs, K. D., & Baumeister, R. F. (2012). What people desire, feel conflicted about, and try to resist in everyday life. *Psychological Science*, *23*, 582–588. https://doi.org/10.1177/0956797612437426

Houben, K., & Wiers, R. W. (2009). Response inhibition moderates the relationship between implicit associations and drinking behavior. *Alcoholism: Clinical and Experimental Research*, *33*, 626–633. https://doi.org/10.1111/j.1530-0277.2008.00877.x

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, *7*, 450–463. https://doi.org/10.1177/1745691612454134

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. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*, e124. https://doi.org/10.1371/journal.pmed.0020124

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/10.1177/0956797610384745

Job, V., & Walton, G. M. (2017). Lay theories of self-control. In C. M. Zedelius, B. C. N. Müller, & J. W. Schooler (Eds.), *The Science of Lay Theories* (pp. 47–69). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-57306-9_3

Johnston, D. W., Allan, J. L., Powell, D. J. H., Jones, M. C., Farquharson, B., Bell, C., & Johnston, M. (2019). Why does work cause fatigue? A real-time investigation of fatigue, and determinants of fatigue in nurses working 12-hour shifts. *Ann. Behav. Med.*, *53*, 551–562.

Kahan, D., Polivy, J., & Herman, C. P. (2003). Conformity and dietary disinhibition: A test of the ego-strength model of self-regulation. *International Journal of Eating Disorders*, *33*, 165–171. https://doi.org/10.1002/eat.10132

Kelley, H. H. (1992). Common-sense psychology and scientific psychology. *Annual Review of Psychology*, *43*, 1–23.

Klein, R. J., Liu, T., Diehl, D., & Robinson, M. D. (2017). The personality-related implications of Stroop performance: Stress-contingent self-control in daily life. *Journal of Research in Personality*, *70*, 156–165. https://doi.org/10.1016/j.jrp.2017.07.006

Koppel, L., Andersson, D., Västfjäll, D., & Tinghög, G. (2019). No effect of ego depletion on risk taking. *Scientific Reports*, *9*, 9724. https://doi.org/10.1038/s41598-019-46103-0

Kruglanski, A. W., Factor, A., & Jaśko, K. (2018). Is "behavior" the problem? *Social Psychological Bulletin*, *13(2)*, e26138. https://doi.org/10.5964/spb.v13i2.26138

Lakens, D. (2013). Using a smartphone to measure heart rate changes during relived happiness and anger. *IEEE Transactions on Affective Computing*, *4*, 238–241.

Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses: Sequential analyses. *European Journal of Social Psychology*, *44*, 701–710. https://doi.org/10.1002/ejsp.2023

Lanaj, K., Johnson, R. E., & Barnes, C. M. (2014). Beginning the workday yet already depleted? Consequences of late-night smartphone use and sleep. *Organizational Behavior and Human Decision Processes*, *124*, 11–23. https://doi.org/10.1016/j.obhdp.2014.01.001

Lanaj, K., Johnson, R. E., & Wang, M. (2016). When lending a hand depletes the will: The daily costs and benefits of helping. *Journal of Applied Psychology*, *101*, 1097–1110. https://doi.org/10.1037/apl0000118

Langer, M. (2018). *Investigating technologically advanced job interview approaches*. [Unpublished dissertation]. Universität des Saarlandes, Germany.

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

Malle, B. F. (2008). Fritz Heider's legacy: Celebrated insights, many of them misunderstood. *Social Psychology*, *39*, 163–173. https://doi.org/10.1027/1864-9335.39.3.163

Martijn, C., Tenbült, P., Merckelbach, H., Dreezens, E., & de Vries, N. K. (2002). Getting a grip on ourselves: Challenging expectancies about loss of energy after self-control. *Social Cognition*, *20*, 441–460. https://doi.org/10.1521/soco.20.6.441.22978

Massin, O. (2017). Towards a definition of efforts. *Motivation Science*, *3*, 230–259. https://doi.org/10.1037/mot0000066

Miketta, S. (2018). *The perseverance of adverse experimental effects after post-experimental debriefings and interventions*. [Unpublished dissertation]. Universität des Saarlandes, Germany.

Milyavskaya, M., Berkman, E. T., & De Ridder, D. T. D. (2019). The many faces of self-control: Tacit assumptions and recommendations to deal with them. *Motivation Science*, *5*, 79–85. https://doi.org/10.1037/mot0000108

Milyavskaya, M., & Inzlicht, M. (2017). What's so great about self-control? Examining the importance of effortful self-control and temptation in predicting real-life depletion and goal attainment. *Social Psychological and Personality Science*, *8*, 603–611. https://doi.org/10.1177/1948550616679237

Mischel, W., & Ebbesen, E. B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, *16*, 329–337. https://doi.org/10.1037/h0029815

Mischel, W., Shoda, Y., & Rodriguez, M. (1989). Delay of gratification in children. *Science*, *244*, 933–938. https://doi.org/10.1126/science.2658056

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., … Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*, 2693–2698. https://doi.org/10.1073/pnas.1010076108

Muraven, M., Collins, R. L., & Nienhaus, 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., Collins, R. L., Shiffman, S., & Paty, J. A. (2005). Daily fluctuations in self-control demands and alcohol intake. *Psychology of Addictive Behaviors*, *19*, 140–147. https://doi.org/10.1037/0893-164X.19.2.140

Muraven, M., Gagné, M., & Rosman, H. (2008). Helpful self-control: Autonomy support, vitality, and depletion. *Journal of Experimental Social Psychology*, *44*, 573–585. https://doi.org/10.1016/j.jesp.2007.10.008

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*, 774–789. https://doi.org/10.1037//0022-3514.74.3.774

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.

Purdue Online Writing Lab (2019). *General Writing FAQs.* Retrieved from https://owl.purdue.edu/owl/general_writing/general_writing_faqs.html

Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three decades of research on motivational intensity theory. In *Advances in Motivation Science* (Vol. 3, pp. 149–186). Elsevier. https://doi.org/10.1016/bs.adms.2016.02.001

Ring, K. (1967). Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*, *3*, 113–123. https://doi.org/10.1016/0022-1031(67)90016-9

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, *7*, 411–426. https://doi.org/10.1177/1745691612454303

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

Sievertsen, H. H., Gino, F., & Piovesan, M. (2016). Cognitive fatigue influences students' performance on standardized tests. *Proceedings of the National Academy of Sciences*, *113*, 2621–2624. https://doi.org/10.1073/pnas.1516947113

Simons, J. S., Wills, T. A., Emery, N. N., & Spelman, P. J. (2016). Keep calm and carry on: Maintaining self-control when intoxicated, upset, or depleted. *Cognition and Emotion*, *30*, 1415–1429. https://doi.org/10.1080/02699931.2015.1069733

Sonnentag, S., Pundt, A., & Venz, L. (2017). Distal and proximal predictors of snacking at work: A daily-survey study. *Journal of Applied Psychology*, *102*, 151.

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

the repliCATS project. (2019). Retrieved from https://replicats.research.unimelb.edu.au/

Tuk, M. A., Zhang, K., & Sweldens, S. (2015). The propagation of self-control: Self-control in one domain simultaneously improves self-control in other domains. *Journal of Experimental Psychology: General*, *144*, 639–654. https://doi.org/10.1037/xge0000065

Tyler, J. M., & Burns, K. C. (2008). After depletion: The replenishment of the self's regulatory resources. *Self and Identity*, *7*, 305–321. https://doi.org/10.1080/15298860701799997

Vadillo, M. A. (in press). Ego depletion may disappear by 2020. *Social Psychology*. https://doi.org/10.1027/1864-9335/a000375

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

Van Reet, J. (2015). Conflict inhibitory control facilitates pretense quality in young preschoolers. *Journal of Cognition and Development*, *16*, 333–350. https://doi.org/10.1080/15248372.2013.833924

von Hippel, W., & Gonsalkorale, K. (2005). "That is bloody revolting!": Inhibitory control of thoughts better left unsaid. *Psychological Science*, *16*, 497–500. https://doi.org/10.1111/j.0956-7976.2005.01563.x

Wright, R. A., Mlynski, C., & Carbajal, I. (2019). Outsiders' thoughts on generating self-regulatory-depletion (fatigue) effects in limited-resource experiments. *Perspectives on Psychological Science*, *14*, 469–480. https://doi.org/10.1177/1745691618815654