



Published in final edited form as:

Soc Psychol Personal Sci. 2021 January 1; 12(1): 14–24. doi:10.1177/1948550619887702.

A Multilab Replication of the Ego Depletion Effect

Junhua Dang¹, Paul Barker², Anna Baumert^{3,4}, Margriet Bentvelzen⁵, Elliot Berkman⁶, Nita Buchholz⁷, Jacek Buczny⁸, Zhansheng Chen⁹, Valeria De Cristofaro¹⁰, Lianne de Vries⁸, Siegfried Dewitte¹¹, Mauro Giacomantonio¹⁰, Ran Gong¹², Maaïke Homan¹³, Roland Imhoff¹⁴, Ismaharif Ismail¹⁵, Lile Jia¹⁵, Thomas Kubiak¹⁶, Florian Lange¹¹, Dan-yang Li¹², Jordan Livingston⁶, Rita Ludwig⁶, Angelo Panno¹⁰, Joshua Pearman⁶, Niklas Rassi¹⁷, Helgi B. Schiöth¹, Manfred Schmitt⁷, A. Timur Sevincer¹⁷, Jiaxin Shi⁹, Angelos Stamos¹¹, Yia-Chin Tan¹⁵, Mario Wenzel¹⁶, Oulmann Zerhouni¹⁸, Li-wei Zhang¹², Yi-jia Zhang¹², Axel Zinkernagel⁷

¹Department of Neuroscience, Uppsala University, Sweden ²Social Cognition Center Cologne, University of Cologne, Germany ³Max Planck Institute for Research on Collective Goods, Bonn, Germany ⁴TUM School of Education, München, Germany ⁵Amsterdam Business School, University of Amsterdam, the Netherlands ⁶Department of Psychology, University of Oregon, Eugene, OR, USA ⁷Department of Psychology, University of Koblenz–Landau, Mainz, Germany ⁸Department of Experimental and Applied Psychology, VU University Amsterdam, the Netherlands ⁹Department of Psychology, The University of Hong Kong, Hong Kong ¹⁰Department of Social and Developmental Psychology, University of Rome “Sapienza,” Italy ¹¹Behavioral Engineering Group, KU Leuven, Belgium ¹²Department of Psychology, Beijing Sport University, China ¹³Amsterdam Institute for Social Science Research, University of Amsterdam, the Netherlands ¹⁴Social and Legal Psychology, Johannes Gutenberg University Mainz, Germany ¹⁵Department of Psychology, National University of Singapore, Singapore ¹⁶Heath Psychology, Johannes Gutenberg University Mainz, Germany ¹⁷Institute of Psychology, University of Hamburg, Germany ¹⁸Laboratoire Parisien de Psychologie Sociale, University Paris Nanterre, France

Abstract

There is an active debate regarding whether the ego depletion effect is real. A recent preregistered experiment with the Stroop task as the depleting task and the antisaccade task as the outcome task found a medium-level effect size. In the current research, we conducted a preregistered multilab replication of that experiment. Data from 12 labs across the globe ($N = 1,775$) revealed a small and significant ego depletion effect, $d = 0.10$. After excluding participants who might have responded randomly during the outcome task, the effect size increased to $d = 0.16$. By adding an informative, unbiased data point to the literature, our findings contribute to clarifying the existence, size, and generality of ego depletion.

Corresponding Author: Junhua Dang, Department of Neuroscience, Uppsala University, Uppsala, Sweden., dangjunhua@gmail.com.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Keywords

ego depletion; self-control; multilab; preregistration

The ego depletion effect refers to the phenomenon that initial exertion of self-control impairs subsequent self-regulatory performance. It has often been tested in between-subject designs using a sequential task paradigm. Participants in such studies complete a task that is thought to require self-control after (a) completing a different self-control task (depletion condition) or (b) completing a control task that does not require self-control (control condition). Participants in the depletion condition generally perform worse than those in the control condition, which is referred to as the ego depletion effect.

According to the strength model of self-control, all acts of self-control draw from a limited mental resource (Baumeister, Vohs, & Tice, 2007). Initial acts of self-control impair subsequent self-control performance because they have consumed the limited resource, leaving insufficient resource for subsequent use. Alternatively, the process model suggests that what underlies ego depletion is not the consumption of some limited resource but the motivated disengagement of work to engage in leisure. That is to say, because effortful control is intrinsically aversive, people generally tend to avoid it. After exerting effortful control, the aversive feeling would accumulate, which leads people not only to more strongly avoid further control but also more strongly value and pursue rewards that can bring gratification (Inzlicht & Schmeichel, 2012).

Regardless of its explanations, observations of the ego depletion effect have had pervasive influence in social psychology and many related research areas (Friese, Loschelder, Gieseler, Frankenbach, & Inzlicht, 2018). The effect has been examined in over 600 independent studies (Carter, Kofler, Forster, & McCullough, 2015) conducted by more than 2,000 researchers (Wolff, Baumann, & Englert, 2018), with various task combinations (for a list of the most frequently used tasks, see Carter et al., 2015; Dang, 2018). The paper that reported the ego depletion effect for the first time (Baumeister, Bratslavsky, Muraven, & Tice, 1998) has been cited over 5,000 times in Google Scholar.

In recent years, however, the ego depletion effect has been challenged on multiple grounds (see Friese et al., 2018, for a review). Meta-analyses indicate that the effect has been severely overestimated due to small-study biases and that the existing evidence may not be sufficient to support the existence of an ego depletion effect (Carter et al., 2015). Perhaps most notably, a multilab replication effort including 23 laboratories ($N = 2,141$) did not find an ego depletion effect that was significantly different from zero, $d = 0.04$, 95% confidence interval (CI) $[-0.07, 0.15]$ (Hagger et al., 2016).

Large-scale replication efforts like the one Hagger and colleagues (2016) have done cannot easily reflect the methodological heterogeneity of ego depletion research (Carter & McCullough, 2018). Rather than being able to operationalize ego depletion in every possible way, they need to select a specific combination of self-control tasks for their study. Based on both conceptual and practical reasons, Hagger and colleagues (2016) chose a letter-crossing task (that differed in self-control demands between the depletion and control condition)

followed by a multisource interference task for their replication project. The appropriateness of this choice has been debated (Baumeister & Vohs, 2016; Dang, 2016; Hagger & Chatzisarantis, 2016) and it remains possible that the ego depletion effect can be observed using task combinations other than the one chosen by Hagger and colleagues. Evidence supporting the existence of an operationalization-specific ego depletion might be difficult to reconcile with the domain-general assumptions of the strength model of self-control. It might, however, help to account for the heterogeneity of previous results and open new avenues for research on self-control processes.

For this reason, we set out to complement the project by Hagger and colleagues with a second multilab preregistered replication study on the ego depletion effect. For this project, we selected a procedure that has been reported to produce a substantial ego depletion effect in a recent preregistered study (Dang, Liu, Liu, & Mao, 2017). Inspired by an up-to-date meta-analysis (Dang, 2018), those authors selected a Stroop task for the first part of the sequential task paradigm, because this task might be particularly effective in affecting subsequent self-control performance. On the subsequent antisaccade task, participants who completed the Stroop task with incongruent stimuli performed significantly worse than participants who completed the Stroop task with congruent stimuli, Hedge's $g = 0.48$. 95% CI [0.18, 0.78]. This finding indicates that the specific task combination used by Dang, Liu, Liu, and Mao (2017) can produce an ego depletion effect. However, on the basis of a single study, it is difficult to conclude whether it does so in a reliable way. Therefore, we preregistered and conducted a high-powered multilab project to replicate the experiment by Dang and colleagues in a standardized way.

Method

Project Organization

This project was initiated by the first author in May 2017. He wrote e-mails to invite researchers who have conducted ego depletion studies. Thirteen labs across the world (4 from Germany, 1 each from the mainland of China, the United States, Belgium, Italy, France, the Netherlands, Austria, Hong Kong, and Singapore, respectively) agreed to participate but the lab from Austria quit at the end of 2017 without collecting any data due to lacking relevant resources. The remaining 12 labs followed the same protocol to collect data independently. All labs have experience of doing ego depletion studies and some of them also participated Hagger et al.'s (2016) replication project.

Preregistered Protocol

The protocol was preregistered in a transparent way in October 2017 (<https://osf.io/4mcnf/>). At the same time, materials including questionnaires measuring individual differences, e-prime scripts with Chinese and English instructions, and description and templates for data extraction have been uploaded. Each non-English and non-Chinese lab translated the questionnaire and e-prime instructions into their local language. Translation efforts were coordinated in the two labs testing Dutch-speaking participants and in the four labs testing German-speaking participants. All data and analytical scripts are available for replicating the results reported here (<https://osf.io/3txav/>).

Experimental procedure—The experimental procedure was the same as that in Dang et al. (2017). Participants first completed a short questionnaire measuring three individual difference variables that might moderate the ego depletion effect: action orientation, lay theory about willpower, and trait self-control. Next, participants engaged in a Stroop task and were randomly assigned to the depletion or the control condition. After the Stroop task, they answered four manipulation check questions before working on the antisaccade task as the outcome measure. Note that three labs measured the individual differences at the end rather than at the beginning of the experiment. The timing of the individual difference measurement was considered as a potential moderator in our auxiliary analyses.

Individual difference measures—Action orientation was measured by the Demand-Related Action Orientation (AOD) subscale of the Action Control Scale (Jostmann & Koole, 2007). The AOD scale consists of 12 items. Each item describes a demanding situation and an action-oriented versus a state-oriented coping way. Participants were asked to indicate the way that best describes their own reaction to that situation. Action-oriented responses were coded as 1 whereas state-oriented responses as 0. Scores summed for the entire scale could range from 0 to 12.

Lay theory of willpower was measured by 6 items developed by Job, Dweck, and Walton (2010). Participants responded on a 6-point scale (1 = *strongly disagree*, 6 = *strongly agree*). Items were scored so that higher values indicate greater agreement with the unlimited-resource theory.

Trait self-control was measured by the 13-item Brief Self-Control Scale (Tangney, Baumeister, & Boone, 2004). Participants indicated the extent to which they agree with each statement on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Higher scores represent better self-control.

Depletion manipulation—Participants assigned to the depletion condition completed a Stroop task in which most trials were incongruent (four different colors, 256 trials, 75% incongruent, such as “BLUE” with red font and “YELLOW” with blue font). In the control condition, all trials were congruent such as BLUE with blue font and YELLOW with yellow font. In each trial, after a 200 ms fixation, the stimulus (i.e., a color word with a specific font color) appeared on the screen until the participant pressed the spacebar, which was then followed by a 500 ms blank screen. Participants were required to read the color of the word aloud and then press the spacebar to proceed toward the next trial.

Manipulation check measures—After the Stroop task, participants answered four manipulation check questions regarding effort (“How much effort did you put into the color-naming task?”), difficulty (“How difficult did you find the color-naming task?”), fatigue (“How tired do you feel after doing the color-naming task?”), and frustration (“Did you feel frustrated while you were doing the color-naming task?”) on a 7-point scale (Hagger et al., 2016).

Outcome measure—Following the manipulation check measures, participants were asked to finish an antisaccade task (Dang, Xiao, Liu, Jiang, & Mao, 2016; Unsworth,

Spillers, Brewer, & McMillan, 2011). Their task was to identify three target letters (B, P, and R) by pressing a corresponding key (the Keys 1, 2, and 3, respectively) as quickly and accurately as possible. At the beginning of each trial, a fixation cross appeared for 200 ms on the screen with a black background. A white “=” was then flashed either to the left or right of the fixation cross for 100 ms, followed by a 50 ms blank screen and a second appearance of the sign “=” for 100 ms at the same location as the first one. This procedure made it appear as though the sign “=” flashed on-screen, which would easily grasp participants’ attention. Following another 50 ms blank screen, the target stimulus (a letter B, P, or R) appeared in the opposite location of the flashing sign for 100 ms, followed by a mask (the letter “H”) for 50 ms and the number “8,” which remained on-screen at the same location as the target stimulus until a response was given. Participants received 30 practice trials (12 practice trials for learning the response mapping and 18 practice trials for doing the formal test) and 120 real trials.

Participants—The sample size was determined to be large enough to detect an effect as large as in the original study ($g = 0.48$) with 80% power (given $\alpha = .05$, two-tailed) in each of the participating labs. Power analysis (G*Power Version 3.1.9.2; Faul, Erdfelder, Lang, & Buchner, 2007) rendered a target sample size of 140 participants (70 in each of the two conditions). Therefore, each lab was required to recruit at least 140 participants who should be between 18 and 30 years, native speakers of the language in which the Stroop color words were presented and not color-blind. They received course credits, money or a gift, or a mix of credits and money or a gift as compensation. The compensation level was determined by local conventions for an experiment lasting for 30–40 min.

There were 1,841 participants in total, of which 1,834 completed the sequential-task paradigm (i.e., the e-prime script). Those whose age was below 18 or above 30 were excluded according to the preregistration protocol, which led to 1,775 participants for final analyses.

Test environment and experimenters—Each participant tested individually in a behavioral lab. The experimenter had to complete the script by himself or herself at least once to get familiar with the script. The experimenter needed to make sure the participant followed the instruction by checking whether the participant’s voice response was consistent with the font color rather than the meaning of the color word during practice trials in the Stroop task in the depletion condition, and whether most feedback provided by the computer were correct during practice trials in the antisaccade task.

Preregistered analyses—According to the preregistration, the data from each lab were first analyzed separately. The primary dependent variable was the error rate of the antisaccade task. The reaction times (RT) were also be examined after trimming (i.e., longer than 2,000 ms and shorter than 200 ms; Unsworth et al., 2011). The t test was used to examine whether there were differences between the two conditions. Also, regression and simple slope analyses were used to test the interaction between the depletion manipulation and each of the three individual difference variables (one by one).

Next, we used meta-analysis to test the overall effect size of the depletion manipulation. We chose the random effects model that allowed to test the homogeneity of the effect sizes across studies by means of the Q statistic (Borenstein, Hedges, Higgins, & Rothstein, 2009). Because the Q statistic cannot be compared across meta-analyses, we also calculated the I^2 , which is expressed as a percentage and is therefore easy to compare, with levels of 25%, 50%, and 75% representing low, medium, and high levels of heterogeneity, respectively. A high level of heterogeneity suggests the majority of the differences observed between individual studies are due to real differences in effect sizes rather than to random sampling errors.

In order to examine the funnel plot asymmetry, we computed two indices: Begg and Mazumdar's (1994) rank correlation (Kendall's τ - b , one-tailed) and Egger's regression test (Egger, Smith, Schneider, & Minder, 1997). The analyses were performed by using the metafor package of R (Viechtbauer, 2010).

To meta-analyze the moderating effects of the individual difference variables, the Liptak–Stouffer Z -score method was used to combine results across labs, weighted by sample size (Karg, Burmeister, Shedden, & Sen, 2011; Whitlock, 2005). First, the p values of the interaction terms were converted to one-tailed p values, with p values less than .50 corresponding to the fact that individuals with action orientation, belief about unlimited resource, and high trait self-control are influenced less by the depletion manipulation, and p values more than .50 corresponding to the fact that individuals with state orientation, belief about limited resource, and lower trait self-control are influenced less by depletion manipulation. Next, these p values were converted to Z -scores using a standard normal curve such that p values less than .50 were assigned positive Z -scores and p values more than .50 were assigned negative Z -scores. Subsequently, these Z -scores were combined by calculating:

$$Z_w = \frac{\sum_{i=1}^k w_i Z_i}{\sqrt{\sum_{i=1}^k w_i^2}},$$

where the weighting factor w_i corresponds to the sample size of each lab, k corresponds to the number of total studies, and Z_i corresponds to the Z -score of each lab. The outcome of this test, Z_w , follows a standard normal distribution and the corresponding probability can be obtained from a standard normal distribution table.

Auxiliary analyses—We also conducted a series of auxiliary analyses to further explore the data. First, we tested whether the estimated effect size of ego depletion was moderated by factors that pertain to participant characteristics and experimental settings. Second, we repeated the preregistered meta-analyses and abovementioned auxiliary analyses after excluding participants who might have responded randomly during the second task. This exclusion criterion was not preregistered, but we regard the corresponding analyses as an informative robustness check.

Results

Preregistered Analyses

Single lab analyses—For the primary dependent variable, the error rate on the antisaccade task, *t* tests showed significantly impaired performance in the depletion condition compared to the control condition in one lab (Baumert, Buchholz, Schmitt, and Zinkernagel), and marginally significant differences in the same direction in two other labs (Bentvelzen, Buczny, de Vries, & Homan; Ismail, Jia, and Tan). Lay theory about willpower interacted with the depletion manipulation to predict the error rate in only one lab (De Cristofaro, Giacomantonio, and Panno), $R^2 = .04$, $p = .014$. Action orientation and trait self-control did not significantly interact with the depletion manipulation in any lab. The descriptive statistics are shown in Table 1.

For the RT, a marginally significant group difference was observed in only one lab (Ismail, Jia, and Tan). Lay theory about willpower interacted with the depletion manipulation to predict the RT in one lab (Kubiak and Wenzel), $R^2 = .03$, $p = .034$. Action orientation and trait self-control did not significantly interact with the depletion manipulation in any lab. The descriptive statistics are shown in Table 2.

Meta-analyses—As shown in Table 3, for the error rate, the weighted average standardized mean difference between the depletion condition and the control condition was significant with a small effect size, $d = 0.10$, 95% CI [0.01, 0.19], $Z = 2.10$, $p = .036$. The *Q* was not significant, $Q(11) = 10.40$, $p = .495$, and the I^2 was 0, which indicates that the effect sizes were rather homogeneous across the participating labs. Neither funnel plot asymmetry index was significant (Kendall's τ -*b* = -0.33 , $p = .153$; Egger's regression coefficient, $t = -0.43$, $p = .680$). The forest plot and the funnel plot are shown in Figures 1 and 2, respectively.

An analogous pattern emerged for the RT. The weighted average standardized mean difference between the depletion condition and the control condition was small and significant, $d = 0.10$, 95% CI [0.00, 0.19], $Z = 2.01$, $p = .044$. The *Q* was not significant, $Q(11) = 5.72$, $p = .891$, and the I^2 was 0. Neither funnel plot asymmetry index was significant (Kendall's τ -*b* = -0.06 , $p = .841$; Egger's regression coefficient, $t = -0.15$, $p = .884$). The forest plot and the funnel plot are shown in Figures 3 and 4, respectively.

With regard to the manipulation check measures, as shown in Table 4, the Stroop version used in the depletion condition was perceived to be more effortful, difficult, and tiring. The effect sizes were homogeneous as indexed by the nonsignificant *Q* and low I^2 . However, on the item regarding frustration, there was no significant difference between the two conditions. Effect sizes were not homogeneous for this measure as indicated by the significant *Q* and medium I^2 .

Regarding the moderating effects of individual difference variables, the Liptak–Stouffer *Z*-score was calculated for each of the three variables (i.e., action orientation, lay theory about willpower, and trait self-control) on both the error rate and the RT. As shown in Table 5, the Liptak–Stouffer *Z*-score method did not reveal any significant moderation of ego depletion

except for lay theory about willpower on the RT, $Z = 2.06$, $p = .039$, suggesting individuals with an unlimited-resource theory were influenced less by the depletion manipulation.

Auxiliary Analyses

Lab-level moderators—We tested whether the estimated effect size from the meta-analysis was moderated by lab-level factors that pertain to participant characteristics and experimental settings. There were eight factors and we tested them one by one on both the error rate and the RT with meta-regression: (1) participants' average age in each lab (continuous variable), (2) the percentage of male participants in each lab (continuous variable), (3) whether the experimenter used a chin fixer (categorical variable: 0 = *no*, 1 = *yes*), (4) whether the experimenter stayed with the participant in the lab (categorical variable: 0 = *no*, 1 = *yes*), (5) the number of experimenters (continuous variable), (6) the percentage of male experimenters (continuous variable), (7) compensation type (categorical variable: 0 = *course credits*, 1 = *money, a gift, or mix of credits and money*), and (8) when the individual difference variables were measured (categorical variable: 0 = *at the beginning*, 1 = *at the end*).

As shown in Table 6, meta-regression showed that there was only one significant moderating effect. The percentage of male participants moderated the ego depletion effect observed on the error rate, $Q_M = 4.00$, $p = .045$, indicating that labs that recruited more male participants tended to find smaller effect sizes.

Exclusion of random responses—Given three possible responses in the antisaccade task, guessing probability was 33%. A Binomial test indicates that, when the error rate is higher than 58.33% (70 errors in 120 trials), the null hypothesis that the participant was responding randomly cannot be rejected at the significance level of 0.05. That is to say, we might be confident to say that a participant performed significantly better than random if he or she had an error rate lower than or equal to 58.33%. Therefore, we repeated the preregistered meta-analyses and the auxiliary analyses (i.e., metaregressions) after excluding participants who completed the antisaccade task at levels indistinguishable from chance (i.e., who were likely to have guessed when required to identify the target letter).

After excluding participants whose error rate was higher than 58.33% ($n = 175$ in the depletion condition and $n = 165$ in the control condition), for the error rate, the weighted average standardized mean difference between the depletion condition and the control condition was still significant, $d = 0.10$, 95% CI [0.00, 0.21], $Z = 1.96$, $p = .049$, as shown in Table 3. The Q was not significant, $Q(11) = 2.89$, $p = .992$, and the I^2 was 0. For the RT, the weighted average standardized mean difference was also significant after exclusion, $d = 0.16$, 95% CI [0.05, 0.26], $Z = 2.93$, $p = .003$. The Q was not significant, $Q(11) = 7.67$, $p = .742$, and the I^2 was 0. After exclusion, the Liptak–Stouffer Z -score method did not reveal any significant moderating effect of the individual difference variables on the ego depletion effect as shown in Table 5. In addition, as shown in Table 6, meta-regression revealed no significant lab-level moderators after excluding chance-level performers.

Discussion

In the current multilab collaborating project, we replicated the ego depletion experiment by Dang and colleagues (2017). Preregistered meta-analyses revealed small and significant ego depletion effects on both response accuracy and latency ($d = 0.10$). After excluding participants who might have responded randomly on the outcome task, the effect size of the accuracy remained the same, but the effect size of the reaction time increased to $d = 0.16$. This effect size estimate was comparable to another recent endeavor in which Garrison, Finley, and Schmeichel (2019) verified the effectiveness of another depleting task implied by Dang's (2018) meta-analysis (i.e., the attention essay task, paired with the Stroop task as the outcome measure in Study 1 and with the attention network test as the outcome measure in Study 2) in two preregistered experiments with over 1,000 participants and reported an effect size $d = 0.20$. Although these results suggest that the ego depletion effect should be real, the effect sizes might be smaller than previously thought. In order to detect an effect size of 0.10, 0.16, and 0.20 with 80% power (given $\alpha = .05$, two-tailed), we will need 1,571, 615, and 394 participants per condition, respectively.

Does this indicate the true effect size of ego depletion is between 0.10 and 0.20? We suggest the answer is either “yes” or “no.” On the one hand, in the current ego depletion literature (as well as Hagger et al.'s and our replications), most studies adopted very brief depletion manipulations that were generally less than 10 min. Participants' responses to these weak manipulations could vary to a great extent, with some participants feeling exhausted while others feeling indifferent or even excited as such tasks may serve to “warm-up” their self-control (e.g., Lopez, Courtney, & Wagner, 2019; Wenzel, Rowland, Zahn, & Kubiak, 2019). Therefore, there is a substantive heterogeneity in the size of ego depletion in the literature and the average effect size is small (e.g., between 0.10 and 0.20). If we continue with these weak manipulations, the answer is more likely to be yes. On the other hand, if stronger manipulations were implemented, the answer should be more likely to be no and the effect size should be increased as the strength of the manipulation increases. Consistently with this proposition, recent evidence showed that the depletion intensity is positively correlated with subsequent fatigue perception (Tsai & Li, 2019). When the manipulation lasted for 1 hr or more, the effect size increased from medium to large (Radel, Gruet, & Barzykowski, 2019; S astad & Baumeister, 2018). Future studies need to systematically test the dose-dependent feature of ego depletion.

In addition, in line with Job et al. (2010), in the full sample, the current project found supporting evidence on the RT for the moderating effect of lay theory about willpower such that people with an unlimited-resource theory were influenced less by the depletion manipulation. Relatedly, a group of studies recently reported reversed ego depletion in individuals who believed that initial exertion is energizing (Savani & Job, 2017). However, after excluding participants who might have responded randomly, which might lead to a more accurate measure of the RT, the moderating effect of lay theory about willpower disappeared. The lack of robustness, in combination with the fact that lay theories did not moderate the effect on error rates, indicates that our results do not provide unambiguous support for a moderating role of lay theories of willpower.

Finally, since our main purpose was to test whether the ego depletion effect is replicable, the current research provides little evidence regarding the mechanisms underlying ego depletion. The observed performance decline after initial exertion may result from insufficient resource for subsequent acts of self-control (Baumeister et al., 2007) or from reduced motivation to exert further control (Inzlicht & Schmeichel, 2012). However, our results pave the way for further theoretical analyses and related empirical examinations because after all it will be meaningless to debate why there is such a phenomenon if the phenomenon itself cannot be observed.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Vetenskapsrådet (2018-06664).

References

- Baumeister RF, Bratslavsky E, Muraven M, & Tice DM (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74, 1252–1265. [PubMed: 9599441]
- Baumeister RF, & Vohs KD (2016). Misguided effort with elusive implications. *Perspectives on Psychological Science*, 11, 574–575. [PubMed: 27474143]
- Baumeister RF, Vohs KD, & Tice DM (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16, 351–355.
- Begg CB, & Mazumdar M (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, 50, 1088–1101. [PubMed: 7786990]
- Borenstein M, Hedges LV, Higgins J, & Rothstein HR (2009). *Introduction to meta-analysis*. Hoboken, NJ: John Wiley.
- Carter EC, Kofler LM, Forster DE, & McCullough ME (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. [PubMed: 26076043]
- Carter EC, & McCullough ME (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.
- Dang J (2016). Commentary: A multilab preregistered replication of the ego-depletion effect. *Frontiers in Psychology*, 7, 1155. [PubMed: 27535004]
- Dang J (2018). An updated meta-analysis of the ego depletion effect. *Psychological Research*, 82, 645–651. [PubMed: 28391367]
- Dang J, Liu Y, Liu X, & Mao L (2017). The ego could be depleted, providing initial exertion is depleting: A pre-registered experiment of ego depletion. *Social Psychology*, 48, 242–245.
- Dang J, Xiao S, Liu Y, Jiang Y, & Mao L (2016). Individual differences in dopamine level modulate the ego depletion effect. *International Journal of Psychophysiology*, 99, 121–124. [PubMed: 26620929]
- Egger M, Smith G, Schneider M, & Minder C (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634. [PubMed: 9310563]
- Faul F, Erdfelder E, Lang AG, & 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. [PubMed: 17695343]
- Friese M, Loschelder DD, Gieseler K, Frankenbach J, & Inzlicht M (2018). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review*. Advance online publication. doi:10.1177/1088868318762183

- Garrison KE, Finley AJ, & Schmeichel BJ (2019). Ego depletion reduces attention control: Evidence from two high-powered preregistered experiments. *Personality and Social Psychology Bulletin*, 45, 728–739. [PubMed: 30239268]
- Hagger MS, & Chatzisarantis NLD (2016). Commentary: Misguided effort with elusive implications, and sifting signal from noise with replication science. *Frontiers in psychology*, 7, 621. [PubMed: 27561304]
- Hagger MS, Chatzisarantis NLD, Alberts H, Anggono CO, Batailler C, Birt AR,...Zwieneberg M (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, 11, 546–573. [PubMed: 27474142]
- Inzlicht M, & Schmeichel BJ (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, 7, 450–463. [PubMed: 26168503]
- Job V, Dweck CS, & Walton GM (2010). Ego depletion—Is it all in your head? Implicit theories about willpower affect selfregulation. *Psychological Science*, 21, 1686–1693. [PubMed: 20876879]
- Jostmann NB, & Koole SL (2007). On the regulation of cognitive control: Action orientation moderates the impact of high demands in Stroop interference tasks. *Journal of Experimental Psychology: General*, 136, 593–609. [PubMed: 17999573]
- Karg K, Burmeister M, Shedden K, & Sen S (2011). The serotonin transporter promoter variant (5-HTTLPR), stress, and depression meta-analysis revisited: Evidence of genetic moderation. *Archives of General Psychiatry*, 68, 444–454. [PubMed: 21199959]
- Lopez RB, Courtney AL, & Wagner DD (2019). Recruitment of cognitive control regions during effortful self-control is associated with altered brain activity in control and reward systems in dieters during subsequent exposure to food commercials. *PeerJ*, 7, e6550. [PubMed: 30842910]
- Radel R, Gruet M, & Barzykowski K (2019). Testing the ego-depletion effect in optimized conditions. *PLoS One*, 14, e0213026. [PubMed: 30845243]
- 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. [PubMed: 28581300]
- Sjstad H, & Baumeister R (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.
- Tangney JP, Baumeister RF, & Boone AL (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271–324. [PubMed: 15016066]
- Tsai MH, & Li NP (2019). Depletion manipulations decrease openness to dissent via increased anger. *British Journal of Psychology*. Advance online publication.
- Unsworth N, Spillers GJ, Brewer GA, & McMillan B (2011). Attention control and the antisaccade task: A response time distribution analysis. *Acta Psychologica*, 137, 90–100. [PubMed: 21470585]
- Viechtbauer W (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.
- Wang Y, Wang L, Cui X, Fang Y, Chen Q, Wang Y, & Qiang Y (2015). Eating on impulse: Implicit attitudes, self-regulatory resources, and trait self-control as determinants of food consumption. *Eating Behaviors*, 19, 144–149. [PubMed: 26402043]
- Wenzel M, Rowland Z, Zahn D, & Kubiak T (2019). Let there be variance: Individual differences in consecutive self-control in a laboratory setting and daily life. *European Journal of Personality*, 33, 468–487.
- Whitlock MC (2005). Combining probability from independent tests: The weighted Z-method is superior to Fisher's approach. *Journal of Evolutionary Biology*, 18, 1368–1373. [PubMed: 16135132]
- Wolff W, Baumann L, & Englert C (2018). Self-reports from behind the scenes: Questionable research practices and rates of replication in ego depletion research. *PLoS One*, 13, e0199554. [PubMed: 29940020]

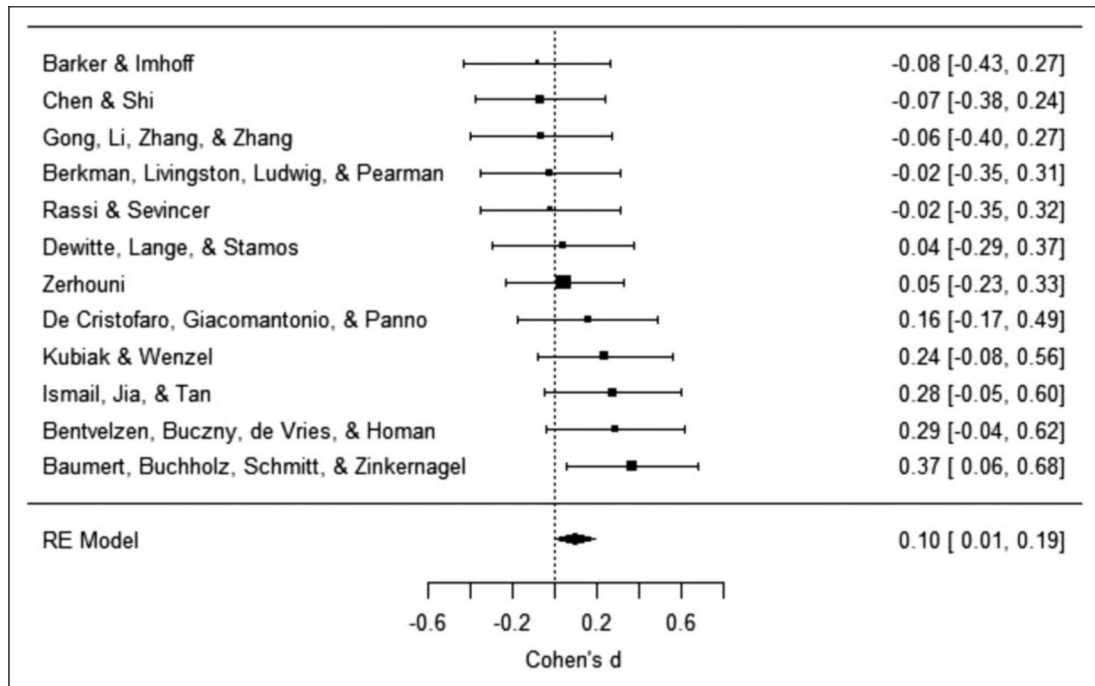


Figure 1.
Forest plot for the error rate.

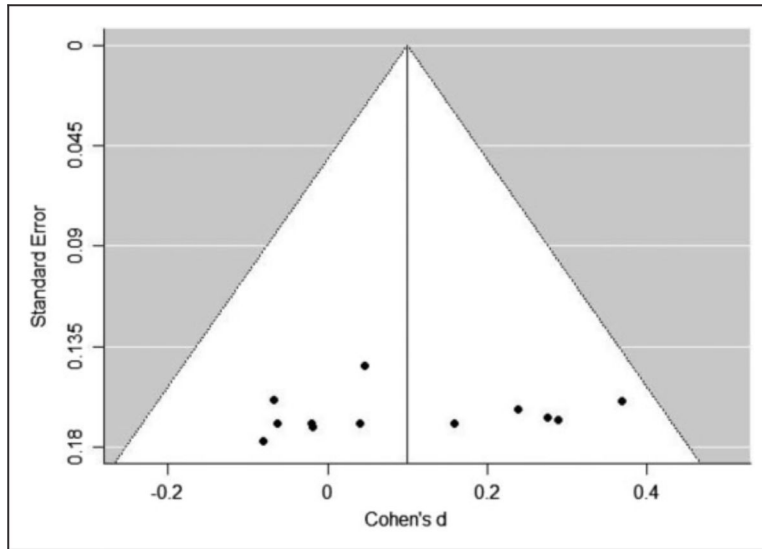


Figure 2.
Funnel plot for the error rate.

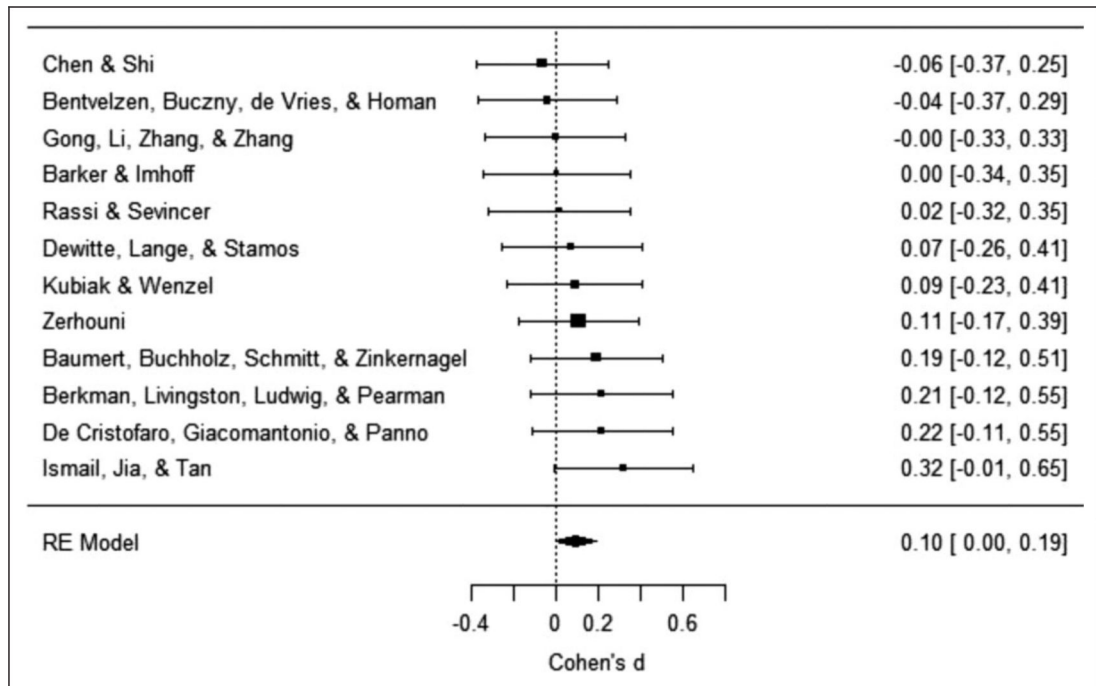


Figure 3.
Forest plot for reaction time.

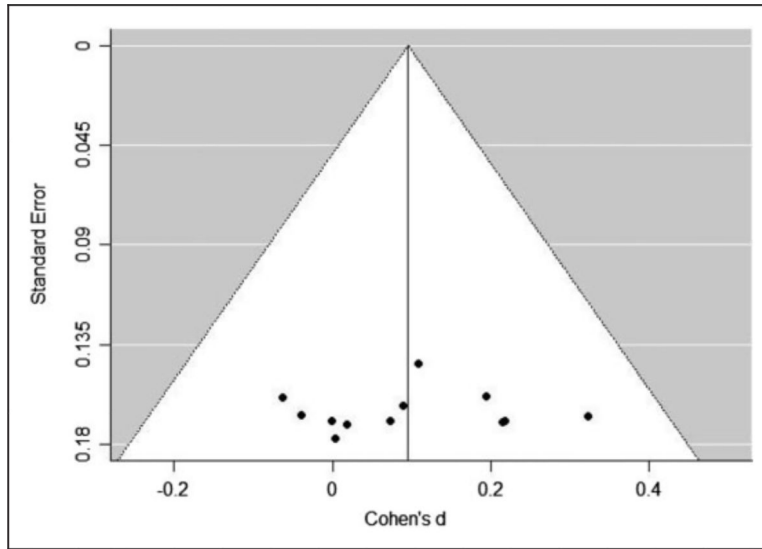


Figure 4.
Funnel plot for reaction time.

Table 1.

The Descriptive Statistics for the Error Rate.^a

Author	$M_{\text{depletion}}$	M_{control}	$SD_{\text{depletion}}$	SD_{control}	t	p
Barker and Imhoff	.55	.56	.13	.11	-0.45	.651
Chen and Shi	.40	.41	.18	.19	-0.42	.672
Gong, Li, Zhang, and Zhang	.39	.40	.16	.19	-0.37	.708
Berkman, Livingston, Ludwig, and Pearman	.23	.23	.16	.16	-0.12	.902
Rassi and Sevincer	.44	.44	.16	.18	-0.11	.910
Dewitte, Lange, and Stamos	.36	.35	.20	.19	0.24	.813
Zerhouni	.40	.39	.20	.19	0.32	.746
De Cristofaro, Giacomantonio, and Panno	.35	.32	.18	.16	0.94	.349
Kubiak and Wenzel	.40	.36	.17	.17	1.47	.144
Ismail, Jia, and Tan	.34	.29	.16	.17	1.66	.099
Bentvelzen, Buczny, de Vries, and Homan	.45	.39	.20	.20	1.73	.086
Baumert, Buchholz, Schmitt, and Zinkernagel	.48	.42	.17	.16	2.33	.021*

^aLabs are ordered by the t value.* $p < .05$.** $p < .01$.*** $p < .001$.

Table 2.

The Descriptive Statistics for the Reaction Times.^a

Author	$M_{\text{depletion}}$	M_{control}	$SD_{\text{depletion}}$	SD_{control}	t	p
Chen and Shi	683.16	694.12	135.22	202.81	-0.40	.690
Bentvelzen, Buczny, de Vries, and Homan	795.93	803.70	206.44	192.50	-0.23	.816
Gong, Li, Zhang, and Zhang	782.71	782.94	193.14	205.01	-0.01	.995
Barker and Imhoff	780.93	780.20	209.73	176.73	0.02	.983
Rassi and Sevincer	801.82	798.48	189.88	174.75	0.11	.915
Dewitte, Lange, and Stamos	764.24	751.76	184.38	156.44	0.43	.668
Kubiak and Wenzel	860.71	843.27	188.42	201.65	0.55	.584
Zerhouni	740.53	722.40	170.29	162.46	0.76	.449
Baumert, Buchholz, Schmitt, and Zinkernagel	826.03	787.80	198.54	194.23	1.23	.221
Berkman, Livingston, Ludwig, and Pearman	639.87	602.11	188.24	161.87	1.27	.207
De Cristofaro, Giacomantonio, and Panno	793.08	753.01	198.37	167.55	1.29	.199
Ismail, Jia, and Tan	696.73	647.74	149.98	153.90	1.94	.054

^aLabs are ordered by the t value.* $p < .05$.** $p < .01$.*** $p < .001$.

Table 3.

Meta-Analytical Effect Size Estimates for the Between-Group Difference in Performance on the Antisaccade Task.

	<u>Full Sample</u>		<u>After Exclusion^a</u>	
	<i>d</i>	95% CI	<i>d</i>	95% CI
Error rate	.10	[0.01, 0.19]	.10	[0.00, 0.21]
Reaction times	.10	[0.00, 0.19]	.16	[0.05, 0.26]

^aResults obtained after excluding participants who might have responded randomly.

*
 $p < .05$.

**
 $p < .01$.

 $p < .001$.

Table 4.

Meta-Analytical Parameters of the Manipulations Check Measures.

	<i>d</i>	95% CI	<i>Q</i>	<i>I</i> ² (%)
Effort	0.50 ^{***}	[0.40, 0.59]	8.57	0
Difficulty	1.16 ^{***}	[1.04, 1.27]	14.59	26
Tiredness	0.26 ^{***}	[0.17, 0.36]	9.28	0
Frustration	-0.02	[-0.15, 0.12]	23.06 [*]	52

* $p < .05$.** $p < .01$.*** $p < .001$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5.

The Liptak-Stouffer Z -Scores for the Moderating Effects of Individual Difference Variables.

	<u>Full Sample</u>		<u>After Exclusion^a</u>	
	Error Rate	RT	Error Rate	RT
Action orientation	0.45	-1.27	0.05	-0.85
Lay theory about willpower	0.65	2.06*	1.30	1.10
Trait self-control	0.55	0.20	0.57	0.70

Note. RT = reaction times.

^aResults obtained after excluding participants who might have responded randomly.

*
 $p < .05$.

**
 $p < .01$.

 $p < .001$.

Table 6.The Q_M Values of Meta-Regression.

	<u>Full Sample</u>		<u>After Exclusion^a</u>	
	Error Rate	RT	Error Rate	RT
Age	0.25	1.05	0.01	0.00
Percentage of male participants	4.00 [*]	1.01	0.01	0.12
Chin fixer use	1.34	0.01	0.04	0.49
Experimenter presence	0.09	0.01	0.06	0.07
Number of experimenters	2.01	1.60	0.96	0.62
Percentage of male experimenter	1.22	0.551	0.12	0.15
Compensation type	0.43	2.63	0.06	0.71
Timing of individual difference measures	0.01	0.03	0.00	0.72

Note. RT = reaction times.

^aResults obtained after excluding participants who might have responded randomly.

*
 $p < .05$.

**
 $p < .01$.

 $p < .001$.