A Multi-Site Preregistered Paradigmatic Test of the Ego Depletion Effect

Kathleen D. Vohs, University of Minnesota

Brandon J. Schmeichel, Texas A&M University

Sophie Lohmann, Max Planck Institute for Demographic Research and University of Illinois at Urbana-Champaign

Quentin F. Gronau, University of Amsterdam

Anna Finley, University of Wisconsin-Madison

…

E.-J. Wagenmakers, University of Amsterdam

Dolores Albarracín, University of Illinois at Urbana-Champaign

(Appendix A contains the entire author list.)

Author note: We thank Roy Baumeister, Eddie Harmon-Jones, Rebecca Schlegel, Uri Simonsohn, Jake Westfall, and Wendy Wood for assistance. This research was supported by a Netherlands Organisation for Scientific Research (NWO) grant to QFG (406.16.528) and to EJW (016.Vici.170.083), as well as an Advanced ERC grant to EJW (743086 UNIFY). Address correspondence to Kathleen Vohs, Carlson School of Management, University of Minnesota, Minneapolis, MN, 55455. kvohs@umn.edu.

Citation: Vohs, K. D., Schmeichel, B. J., Lohmann, S., Gronau, Q., Finley, A. J.… Wagenmakers, E.-J., & Albarracín, D. (in press). A multi-site preregistered paradigmatic test of the ego depletion effect. *Psychological Science.*

Abstract

We conducted a preregistered, multi-laboratory project (k = 36; N = 3531) to assess the size and robustness of ego depletion effects using a novel replication method, termed the paradigmatic replication approach. Laboratories implemented one of two procedures that intended to manipulate self-control and tested performance on a subsequent measure of self-control. Confirmatory tests found a non-significant result, d = 0.06. Confirmatory Bayesian meta-analyses using an informed prior hypothesis (δ = 0.30; *SD* = 0.15) found the data were four times more likely under the null than the alternative hypothesis. Hence, preregistered analyses did not find evidence for a depletion effect. Exploratory analyses on the full sample (i.e., ignoring preregistered exclusion criteria; see supplemental online materials) found a statistically significant effect (d = 0.08), with data about equally likely under the null and informed prior hypotheses. Exploratory moderator tests suggested that the depletion effect was larger for participants reporting more fatigue but was not moderated by trait self-control, willpower beliefs, or action orientation.

A Multi-Site Preregistered Paradigmatic Test of the Ego Depletion Effect

The theory of ego depletion was introduced in 1998 and quickly gained interest from scholars and lay audiences alike. Ego depletion is a theory of how self-control operates, with self-control defined as the capacity to alter a predominant response tendency, control impulses, and engage in volitional behavior. The central notion is that self-control operates like a limited resource, such that using self-control on an initial task renders subsequent self-control less successful than if not deployed earlier (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Muraven, Tice, & Baumeister, 1998).

The concept of ego depletion has been widely influential. The seminal article (Baumeister et al., 1998) has had “transformational” impact (Nosek et al., 2010, Supplement). In addition to a multitude of empirical articles, the theory inspired multiple new theories as well (e.g., Evans, Boggero, & Segerstrom, 2015; Inzlicht & Schmeichel, 2012; Job, Dweck, & Walton, 2010; see Baumeister & Vohs, 2016a, for a review). In short, the theory has been highly generative, both empirically and theoretically.

In recent years the evidentiary basis of ego depletion has been challenged and, in response, we embarked on a multi-site, preregistered test of the phenomenon. Challenges to ego depletion have come in two main forms: Meta-analytic analyses (Carter et al., 2015) and a multi-site registered replication study (Hagger et al., 2016). Those investigations cast doubt on ego depletion theory but have been criticized on methodological and analytical grounds (Baumeister & Vohs, 2016b; Garrison, Finley, & Schmeichel, 2019; Friese et al., 2019; Inzlicht, Gervais, & Berkman, 2015). Germane to the current study is that the previous replication study used methods uncommon to ego depletion studies (see Hagger & Chatzisarantis, 2016, for a rebuttal). As a result, we conducted a multi-site, preregistered study with methods more common to the literature and more paradigmatic of the construct.

Paradigmatic Replication Approach

The current approach tested a hypothesis derived from the theory of ego depletion and aimed to create a new model for replication studies. Termed the *paradigmatic replication approach*, it made multiple changes to existing models (see Spellman & Kahneman, 2018, for how the current project differs from others). Chiefly and briefly, the procedures did not draw from any one published study. Instead, candidate procedures were selected for how well they represent the phenomenon — hence the paradigmatic moniker. Table 1 outlines key elements of the Paradigmatic Replication Approach.

Additionally, the paradigmatic approach involved crowdsourcing with experts in depletion research, scholars who sought to participate in data collection, and statistical advisors. Experts generated possible tasks for the study's procedures, focusing on their paradigmatic fit with the construct. Labs then vetted those tasks for whether they would provide good tests of the hypothesis and could be executed in their laboratories.

We recruited a group of scholars with little or no prior connection to ego depletion research to serve as an Advisory Board. They made recommendations on data analytic models, data analysis procedures, and study preregistrations. Prior to data collection, the lead author (KV) created instructional videos for participating laboratories depicting mock experimental sessions and held virtual meetings with experimenters to answer questions. After completing data collection, laboratories sent their data to a handler who created a master dataset and blinded the data, which was then sent to the data analysis team. The analysis team conducted preregistered analyses before sharing results with the lead authors (KV and BS), who then generated recommendations for exploratory analyses. Lead authors had access to the data only after the analysts had done their work (Table 1).

Experimental Protocols

Laboratories used one of two protocols. (The term *protocol* refers to each combination of independent and dependent variables.) The E-task protocol used a manipulation that varied instructions to cross out the letter “e” within printed text and measured subsequent self-control by persistence on unsolvable geometric puzzles. Both tasks are common in the published depletion literature (e.g., Baumeister et al., 1998; DeWall et al., 2007; Vohs et al., 2008). The writing task protocol used a manipulation that had people write a story with or without difficult instructions and a self-control outcome measure involving answering questions that benefited from controlled cognitive processing. The Cognitive Estimation Test (CET; Bullard et al., 2004; Fein et al., 1998) is thought to require self-control because answers cannot be determined algorithmically or with declarative knowledge. These tasks also have been used in the depletion literature (e.g., Mead et al., 2009; Schmeichel, 2007; Schmeichel, Vohs, & Baumeister, 2003).

The primary hypothesis concerned ego depletion. In line with the theory, we expected that people randomly assigned to use self-control during an initial task would show worse self-control subsequently, compared to people who did not use self-control initially. We expected the magnitude of the effect to be equivalent across protocols (see preregistration,

https://osf.io/952mv/?view_only=a81b3b1fd3e64898832cf19648b8c1dd).

We chose manipulation checks common to the depletion literature, namely participants' reports of the difficulty of the initial task, degree of effort required for it, and feelings of frustration from it (Hagger, Wood, Stiff, & Chatzisarantis, 2010). Other self-report measures included reports of being tired or fatigued. We predicted that compared to the non-depletion condition, people in the depletion condition would report that the initial task was more effortful and difficult—this was the primary manipulation check. We also expected the manipulation to make them feel more tired, fatigued, and frustrated. Additionally, Inzlicht and Schmeichel (2012) proposed that depletion hampers motivation, which we tested with self-reports of being motivated and wanting to do well on the outcome task. Inzlicht and Schmeichel's theory would predict lower motivation among people in the depletion, compared to non-depletion, condition. The original ego depletion model does not make this prediction and thus anticipates no differences in motivation.

We tested potential moderator variables, both by states thought to arise from the manipulations and trait measures. On the former, we tested moderation by manipulation check responses, predicting that being in the depletion condition and reporting higher scores on those items would result in larger depletion effects.[1] The more effortful, fatiguing, or frustrating the initial task, the more it should undermine subsequent self-control performance (e.g., Clarkson et al., 2010; Dang, 2016).

[1] The term *depletion effect* refers to lower performance on outcome tasks among participants who had previously exerted self-control. The term *depletion condition* refers to an initial task designed to require self-control whereas *non-depletion condition* refers to a task designed to require relatively less self-control.

We tested potential moderation by individual differences as well. We measured beliefs about willpower (Job et al., 2010), decision-related action orientation (Kuhl, 1994), and trait self-control (Tangney, Baumeister, & Boone, 2004). Each has been found to moderate depletion effects in prior research. We predicted that people who believe that willpower is a limited resource (Job et al., 2010) or are less inclined toward action orientation (Jostmann & Koole, 2007) would show stronger depletion effects. Findings on trait self-control are mixed, with stronger depletion effects found among people possessing higher (e.g., Dvorak & Simons, 2009) and lower trait self-control (e.g., DeWall et al., 2007), therefore we registered a research question with no firm predictions regarding trait self-control.

Other project features aimed to track potential moderation variables. To assess differences in study execution, laboratories provided videos of experimenters, which were subjected to independent ratings. Other potential moderators included the number of publications by laboratories' principal investigators (PIs), number of depletion studies published by the PI, and laboratory location (see Supplementary Online Materials [SOM]).

The study also collected demographic information. Demographic variables included gender identification (response options: female, male, other), age, and language spoken at home.

Methods

Participants

Thirty-six laboratories (see Appendix B) tested 3531 people (2375 women, 1130 men, 11 listed "other," and 15 did not report gender: *M* age = 20.92, *SD* = 5.19). Most of the laboratories were located in the U.S. (*k* = 23), plus five labs in Germany, three in Canada, two in the Netherlands, two in Australia, and one in Italy. Sixteen laboratories chose to use the writing task protocol (*n* = 1679) and 20 laboratories chose the E-task protocol (*n* = 1852). Among all participants, 1762 were randomly assigned to the depletion condition and 1769 were randomly assigned to the non-depletion condition. Based on preregistered criteria, we excluded 30.25% (*n* = 1068) of all participants in confirmatory data analyses, most often because of excessive errors on the E-task, not being a native speaker of the laboratory's language, or failing to comply with instructions to not use their phone (for more information on exclusions and how this rate compares to other multi-site replications, see Table 2 and SOM). The exclusion rate exceeded our informal expectations and prompted exploratory analyses on the full sample of participants (i.e., with no exclusions), which are reported in the SOM.

Protocol Generation and Creation

Two months prior to the start of data collection, a list of possible operationalizations of the independent and dependent variables was generated by experts in depletion research and sent to scholars who had indicated interest in participating in this project. Those scholars provided feedback on each of the operationalizations as to how effective they believed the tasks would be for testing ego depletion and how feasible they would be to conduct.

For potential manipulation tasks, effectiveness was defined as the extent to which the task would be depleting for their participants. For potential outcome tasks, the effectiveness item asked the extent to which the task would yield enough variance within their sample so that a depletion effect could be detected.

Analyses identified the top-rated procedures, leading to three protocols. Participating labs then ranked their preferences as to which protocol to execute. As it turned out, all laboratories save for two chose either the E-task protocol or writing task protocol; we assigned those laboratories to their second choice. The two tasks used as manipulations and the two tasks used as outcome measures received the top combined ratings of effectiveness and feasibility.

Prior to data collection, laboratories received training on how to execute each protocol via video tutorials and virtual meetings. Methods, predictions, exclusion criteria, and analytical specifications were preregistered prior to data analysis (https://osf.io/952mv/?view_only=a81b3b1fd3e64898832cf19648b8c1dd). The SOM contains additional methods details.

Experimental Procedures

Overview. Both protocols followed the same basic procedure, with the only difference being the operationalization of the independent and dependent variables.

Participants were told that the study examined different types of cognitive processes and specifically people's responses to tasks that tap into different cognitive processes. They completed the independent and dependent variable tasks, which varied by protocol. Next, they completed manipulation checks, motivation reports, individual differences scales, demographic questions, and a post-experimental questionnaire (https://osf.io/952mv/?view_only=a81b3b1fd3e64898832cf19648b8c1dd).

E-task protocol. First, participants completed a task that involved crossing off all instances of the letter E on a sheet of text, after which everyone received a new page of text. Depending on experimental condition, participants either followed the same rules as before and crossed out all instances of the Es (non-depletion condition) or were given new rules requiring them to selectively cross out Es as a function of whether there was a vowel before or after the letter (depletion condition). The task had time limits: 7 minutes for the first page and 8 minutes for the second.

The experimenter then introduced the dependent measure—a figure tracing task, which was described as a spatial abilities task. The task involved using a highlighter marker to trace each figure in its entirety without picking up the highlighter or crossing over the same line segment twice. Once assured that participants understood, experimenters laid down stacks of the three test images, telling participants they could quit the task anytime by ringing a bell on their desk. Unbeknownst to participants, two of the three figures could not be traced as instructed (i.e., they were unsolvable). Experimenters started timing after leaving the room and stopped timing when participants indicated they were done with the task (or after 20 minutes).

Time spent on the task (i.e., duration) and number of sheets attempted formed the dependent measure of self-control. Number of figure tracing sheets used (representing attempts) and duration of the task were standardized separately and added to create an overall figure tracing score (r = 0.39, 95% CI [0.35, 0.43]).

Writing task protocol. Participants’ first task was to write a story about a recent trip. Participants in the non-depletion condition received no additional instructions. Participants in the depletion condition were further instructed not to use words containing the letters *A* or *N* in their story. Both conditions wrote for 5 minutes. After the writing task, the experimenter introduced the dependent measure, the Cognitive Estimation Test (i.e., CET; sample item: “How many seeds are there in a watermelon?”). Participants were told that they should give their best guess on each item. There was no time limit on the CET.

CET responses were awarded points for degree of accuracy (0-2) in accordance with published standards (Bullard et al., 2004; Fein et al., 1998). After determining the number of valid responses given by each participant (SOM), points were averaged to form a final CET score, which then was standardized.

Manipulation checks. After the dependent measure, participants in both protocols completed manipulation check items and other task-related reports. They reported the difficulty and effort required for the manipulation task, which were the key manipulation check items. Participants also reported how much the manipulation task made them feel frustrated, fatigued, and tired. Two additional items assessed participants' motivation for the dependent measure. They reported how motivated they felt during the task and how much they wanted to do well on it. All items were rated on Likert scales from *1 = not at all* to *7 = very*.

Individual differences. Trait measures were administered last. Items were averaged to create composite scores.

Participants completed the 12-item Decision-Related Action Orientation subscale of the HAKEMP (Kuhl, 1994), which measures whether people take action to work on tasks or tend to put them off (*M* = 5.78*; SD*= 2.85; α = 0.71). Sample item: "When I know I must finish something soon: A) I have to push myself to get started, or B) I find it easy to get it done and over with" (participants receive 1 point for each action-orientated option they chose). Next, they completed the 13-item Trait Self-Control Scale (Tangney et al., 2004), which measures dispositional self-control tendencies (*M* = 3.23; *SD* = 0.63; α = 0.81). Sample item: "I am good at resisting temptation" (1 = *not at all like me*; 5 = *very much like me*). Last, participants completed the 6-item Strenuous Mental Activity subscale of the Implicit Theories about Willpower Scale[2] (Job et al., 2010), which measures whether people think that self-control is a limited resource (*M* = 4.18, *SD* = 0.90; α = 0.84; *n* = 2452). Sample item: “After a strenuous mental activity, your energy is depleted and you must rest to get it refueled again” (1 = *strongly agree*; 6 = *strongly disagree*; scores were reversed such that higher numbers indicated stronger beliefs that self-control is a limited resource).

Data and Analytic Procedures

Advisory board. We formed a methodological and statistical Advisory Board. Members were selected for being experts in open data, replications, or statistical techniques (i.e., frequentist and Bayesian meta-analyses).[3] Advisory Board members provided invaluable help in formulating hypotheses, suggesting analytical models, analyzing data, and preregistering the project.

Dataset procedures. After labs completed data collection, they sent a dataset to a member of the organizing team who previously had been uninvolved in depletion research. This scholar’s role was to receive, merge, and otherwise handle the data, thereby ensuring that the lead authors (KV and BS) would not have access to the data until after the analysts[4] from the Advisory Board performed analyses.

Two steps were taken to ensure data integrity. One involved blinding the data prior to analyses. The data handler switched the names of the columns containing the main dependent measures with another column before passing the dataset off to the

[2] Due to formatting errors, some laboratories omitted the Implicit Theories of Willpower Scale, resulting in different sample sizes.

[3] Advisory Board members were Dolores Albarracín, Will Gervais, Quentin Gronau, Sophie Lohmann, EJ Wagenmakers, Jake Westfall, and Wendy Wood.

[4] Dolores Albarracín, Quentin Gronau, Sophie Lohmann, and EJ Wagenmakers

analysts. Thus, lead authors did not have access to the data until after the analysts did, nor did they conduct analyses. After initial analyses were conducted, the dataset was unblinded. As a second step the analysts conducted all of the hypothesis tests and populated the data displays.

Frequentist statistics. Prior to excluding participants according to preregistered criteria, we standardized all outcome variables and centered all continuous moderators for ease of interpretation. For the frequentist approach, we conducted random-effects (RE) meta-analyses on each laboratory's Cohen's *d* effect size, representing the difference between the non-depletion and depletion conditions. (Fixed-effects [FE] analyses are reported in parentheses.) Larger effect sizes indicate a stronger ego depletion effect (i.e., lower scores on the dependent measures of self-control). Analyses were conducted in R (R Core Team, 2019). Moderators were tested using multi-level linear models in the individual-level analyses (Bates et al., 2015) and using random-effects meta-regression for meta-analytic analyses at the lab level (Viechtbauer, 2010).

Bayesian statistics. Bayes factors addressed the evidentiary basis of the depletion effect. To address the question, "Does the effect exist?" we pitted a point-null hypothesis, which states that the effect is absent, against an informed one-sided alternative hypothesis centered on depletion effect of $\delta = 0.30$ with a standard deviation of 0.15. The preregistered alternative hypothesis estimate was based on effect sizes from two prior large-scale depletion investigations: Hagger et al.'s (2010) meta-analysis, which reported an overall effect size of $d = 0.62$, and Hagger et al.'s (2016) registered replication report, which reported an overall effect size of $d = 0.04$. We split the difference and arrived at $\delta = 0.30$ ($SD = 0.15$). In line with the one-sided nature of the depletion hypothesis, the prior was truncated at zero to allow only positive effect size values. We computed Bayes factors (e.g., Jeffreys, 1939) to quantify the relative support for the informed ego depletion versus the point-null hypothesis.

Subsequent analyses provided information on the size of the ego depletion effect after having seen the data. Posterior distributions for the effect size addressed the question “Assuming that there is an effect, how large is it?”

We conducted a Bayesian meta-analysis on the *t*-test of the depletion effect from each laboratory. In contrast to the classical approach, this approach used Bayesian model averaging, which combines the results of fixed- and random-effect models according to their plausibility given the data (Gronau et al., 2017; Scheibehenne, Gronau, Jamil, & Wagenmakers, 2017). We quantified the model-averaged evidence for an effect and identified a model-averaged posterior distribution for the meta-analytic effect size. For this meta-analysis, we specified the informed prior for effect size and a prior distribution for between-study heterogeneity. We used a preregistered informed Beta (1,2) distribution for the between-study standard deviation (van Erp, Verhagen, Grasman, & Wagenmakers, 2017).

Results

The results section reports preregistered and thus confirmatory analyses on the reduced sample (i.e., after excluding participants on the basis of preregistered criteria; Table 2). First, we report results on the manipulation check items using both frequentist and Bayesian approaches. Next are tests of whether the depletion manipulations affected subsequent self-control using both frequentist and Bayesian approaches. This section is followed by frequentist statistical tests of proposed moderator variables. (Bayesian analyses were not available for moderator tests.)

Results are presented such that higher numbers indicate results in line with hypotheses. That is, for the manipulation checks, higher numbers indicate that depletion condition participants reported stronger feelings than did non-depletion participants. For the main hypothesis-testing results, higher numbers indicate worse performance on the outcome task in the depletion (versus non-depletion) condition, which is taken as evidence of a depletion effect.

Exploratory tests can be found in the SOM. They include manipulation checks, hypothesis tests using both Bayesian and frequentist approaches, and moderation analyses. Most of the exploratory analyses are on the full sample (that is, without excluding any participants).

Manipulation Checks

Frequentist analyses. Meta-analyses were conducted to check the effectiveness of the depletion task (Table 3). Ratings of how much effort the manipulation task required and its difficulty formed an internally consistent scale (Spearman-Brown coefficient = .79) and therefore were averaged into a single index of effort; we preregistered the effort index as the primary manipulation check. As predicted, participants in the depletion condition reported that the manipulation task was more difficult and effortful than did participants in the non-depletion condition. Although scores on the effort index showed substantial heterogeneity across laboratories, with effect sizes ranging from *d* = 0.08, 95% CI [-0.65, 0.81] to *d* = 4.57, 95% CI [3.21, 5.94], there was evidence that the manipulation worked as intended.

We tested whether scores on the effort index differed by protocol, coded such that the intercept (*d* = 1.76, 95% [1.66, 1.86], I^2 = 0%) represents the average effect across both protocols (-.5 = E-task; .5 = Writing task). We did not expect protocol to moderate scores on the effort index but preregistered that we would test each protocol separately if protocol were a significant moderator—which it was (*b* = 2.61, 95% CI [2.41, 2.81]). Therefore, we calculated planned contrasts to examine the effect separately for each protocol. The depletion task was rated as more difficult and effortful than the non-depletion task in both protocols, but the difference was larger in the writing task protocol (*d* = 3.09, 95% CI [2.87, 3.30], I^2 = 39.29%) than in the E-task protocol (*d* = 0.46, 95% CI [0.34, 0.57], I^2 = 0%). These results suggest that the depletion manipulation was more effortful than the non-depletion manipulation, as intended, and that one protocol was more effortful than the other.

Reports of how tired and fatigued participants felt after performing the manipulation task were internally consistent (Spearman-Brown = .90) and, as preregistered, were averaged to form an index of fatigue. As predicted, the main effect of depletion condition was significant, such that participants reported more fatigue in the depletion than the non-depletion condition. Also as expected, participants in the depletion condition reported feeling more frustrated than did participants in the non-depletion condition (Table 3; also Tables S4 and S5).

Reports of motivation and wanting to do well on the dependent measure formed an internally consistent scale (Spearman-Brown = .74). The two items were standardized and averaged to form a motivation index. We preregistered competing predictions: (a) that there will be no depletion condition effect (in line with the ego depletion theory) or (b) that depletion condition participants would report lower motivation than non-depletion participants (in line with Inzlicht & Schmeichel, 2012). Consistent with (a), there was no difference in self-reported motivation (Table 3; Tables S4 and S5).

Bayesian analyses. To quantify the predictions under H1, a model-averaged Bayesian meta-analysis using a one-sided Cauchy prior on effect size μ with mode 0 and scale 0.707 was conducted. Given that the preregistration plans for the primary outcome variable specified using a Beta(1,2) prior distribution for the between-study heterogeneity τ, we adopted that approach here. However, in work succeeding the preregistration we consistently have used an inverse-Gamma prior with shape 1 and scale 0.15 (e.g., Gronau et al., 2017; van Erp et al., 2017), which we used here as well. Hence below we report the results both for the Beta prior and for the inverse-Gamma prior. Noticeable differences between these priors are due to the fact that the Beta prior does not allow values for τ higher than 1, contrary to what the data suggested.

For the effort index, BF(Beta prior) > 1.797693e+308 and BF(inverse-Gamma prior) = 1,123,563; for feelings of frustration, BF(Beta prior) = 2,727,844,064 and BF(inverse-Gamma prior) = 85,152; and for the fatigued index, BF(Beta prior) = 5.68 and BF(inverse-Gamma prior) = 6.13. For motivation, both priors yielded the same Bayes factor in favor of the null hypothesis, BF_{+0} = 0.029 (in other words, BF = 34.48 in favor of the null). These results provide clear evidence that, overall, the depletion manipulations increased feelings of effort and frustration and moderate evidence that depletion increased feelings of fatigue. As for self-reported motivation, we found that the depletion manipulations did not affect it.

Performance on the Outcome Tasks: Hypothesis Test Analyses

Frequentist analyses. Contrary to predictions, meta-analytic results showed that the standardized mean performance difference between the depletion and the non-depletion conditions was not statistically significant, *d* = 0.06 [-.02, 0.14] (see Table 4; Figure 1).

Bayesian analyses. The presence of a depletion effect was then tested using Bayesian analyses. In these analyses, a Bayes factor of BF_{+0} = 10 would indicate that the data are 10 times more likely under the informed alternative hypothesis, which is centered on δ = 0.30, than under the point-null hypothesis. Correspondingly, a BF_{+0} = 1/10 would indicate that the data are 10 times more likely under the point-null hypothesis than under the informed alternative hypothesis.

The meta-analytic Bayes factors quantify the overall evidence in favor of either the informed alternative hypothesis or the point-null hypothesis across all laboratories simultaneously. The meta-analytic Bayes factor of focal interest is the model-averaged one (Figure 2). For comparison, we displayed the meta-analytic Bayes factor for the fixed- and random-effect models separately. All three meta-analytic Bayes factors showed close agreement and favored the point-null hypothesis to approximately the same degree (Figure 2). The model-averaged Bayes factor indicated that the data are 4.4 times more likely under the point-null hypothesis (which states that the effect is absent) than under the one-sided informed alternative hypothesis of a depletion effect

(Figure 3).[5] This Bayes factor value indicates moderate evidence in favor of the point-null hypothesis according to the classification scheme proposed by Jeffreys (1939).

Posterior distributions. All posterior distributions supported only positive effect size values, which follows from an *a priori* decision to use an informed prior that does not allow negative effect size values. When examining individual laboratories' data, many showed a shift toward updating the effect size toward zero, indicating that even if the effect was not zero, it was likely smaller than the expected *d* = 0.30.

Assuming a nonzero effect, an inspection of the data across the individual laboratories did not permit strong conclusions about the size of the effect because of the large uncertainty associated with individual laboratories' effect sizes. To account for findings from all laboratories simultaneously, we considered the results of the model-averaged meta-analysis. We concluded that the data have shifted our beliefs about the effect size of ego depletion from one centered around δ = 0.30 toward zero. The posterior median was 0.08, 95% CI [0.01, 0.16] (Figure 3).

Potential Moderators

Protocol type. We first checked whether outcomes varied by protocol (the specific combination of manipulation and dependent measures). The dependent measure was performance, and protocol type was contrast-coded (-0.5: E-task, 0.5: Writing task) so that the intercept represented the average effect across both protocols. A meta-analytic test (main effect random-effects [RE] model: d = 0.06, 95% CI [-0.02, 0.14], moderator b = -0.10, 95% CI [-0.26, 0.06]) indicated that protocol type was not a

[5] Our recent work uses an inverse-Gamma distribution, which we applied to the confirmatory depletion hypothesis test. Results did not appreciably change compared to those using the Beta distribution (Figure 4). Using an inverse-Gamma prior for between-study heterogeneity tau, the model-averaged meta-analytic BF_{+0} = 0.228 or, expressed in favor of the null, BF_{0+} = 4.39.

significant moderator, suggesting that the magnitude of the effect did not differ across protocols.

The total score on the figure tracing task was the combination of the number of puzzle sheets participants used (as an indicator of attempts) and time spent on the task. For the combined measure of figure tracing duration and attempts in the E-task protocol, we found a non-significant effect of condition (Table 4).

We preregistered our intention to examine separately the two components of the E-task protocol's performance outcome (i.e., the figure tracing task). In prior work, the two components correlated highly and showed parallel effects (e.g., Fennis, Janssen, & Vohs, 2009; Vohs et al., 2008). In the current data, however, the two figure tracing components exhibited only a moderate correlation. *r* = .39, 95% CI [0.35, 0.43].

Examining the two components separately, the effect of condition on number of attempts was not statistically significant (unstandardized descriptives: non-depletion condition *M* = 19.87, *SD* = 9.92; depletion condition *M* = 19.36, *SD* = 10.41; Table 4).

In contrast, there was a significant effect on duration (RE: *d* = 0.15, 95% CI [0.02, 0.29]; fixed-effects [FE] model: *d* = 0.13, 95% CI [0.02, 0.25]; Table 4). Participants in the depletion condition gave up about 27s sooner on the figure tracing task than participants in the non-depletion condition (unstandardized descriptives: non-depletion condition *M* = 1012.20s, *SD* 266.30; depletion condition *M* = 985.10s, *SD* = 283.52).

We preregistered additional moderation tests of manipulation check ratings and individual differences. We did not, however, specify the statistical approach we would use, so we refer to them as exploratory analyses and report them in the SOM. The results showed that the only variable to act as a significant moderator was the self-reported index of fatigue. Performance was worse in the depletion (compared to non-depletion) condition among participants who reported being more fatigued by the manipulation task (Table S2, Figure S4).

Discussion

We tested an ego depletion hypothesis on more than 3500 participants in 36 independent laboratories, which used one of two experimental protocols. The results lead us to conclude that depletion is not as reliable or robust as previously assumed.[6] Confirmatory frequentist analyses indicated that the two conditions did not differ, although outcome performance was directionally worse in the depletion condition compared to the non-depletion condition (*d* = 0.06; Table 4). Confirmatory Bayesian tests found more evidence for the absence than presence of an ego depletion effect (Figure 3). Hence, preregistered analyses did not show a depletion effect.

Our preregistered exclusion criteria led us to exclude data from nearly a third of the overall sample, which exceeded expectations. Frequentist exploratory analyses using the full sample of participants (without exclusions) found a statistically significant but small (*d* = 0.08; Table S1, Figure S1) depletion effect. Comparable Bayesian analyses showed no clear evidence to support or refute the informed alternative hypothesis in support of a depletion effect (Figure S2, Figure S3).

Moving back to frequentist tests, the findings suggested that self-reported fatigue acted as a moderator. The more that depletion condition participants felt fatigued, the worse their subsequent self-control (Table S2, Figure S4). This pattern is congruent with prior evidence regarding the role of subjective fatigue in the ego depletion effect (e.g.,

[6] This language reflects our preregistered conclusion if analyses showed a non-significant result.

Clarkson et al., 2010). There was no statistically significant moderation by self-reported effort, frustration, or motivation. We also tested a host of plausible trait moderators that evinced little predictive value.

Interpretations, Implications, and Integrations

How do these findings inform an understanding of ego depletion? We see several potential interpretations of these findings. One is that there is no depletion effect. The preregistered analyses support this interpretation (Table 4, Figures 1 and 2).

A second perspective is that the reliability of the effect is still unknown, supported by the inconclusive exploratory Bayesian results on the full sample. Both the null hypothesis and informed alternative hypothesis, specifying a 70% probability that the effect size falls between $\delta = 0.15$ and $\delta = 0.45$, fit the full-sample data about equally well (Figures S2 and S3).

A third perspective is that there may be a reliable, but small, depletion effect. The exploratory frequentist analyses on the full sample support this interpretation (Table S1, Figure S1). Exploratory analyses showing significant moderation by self-reported fatigue further suggest that depletion effects may be conditional (Table S2, Figure S4).

There are several implications of these views for future research. First, some analyses hinted at a small depletion effect, but those were exploratory analyses and hence confidence in them should be low until they are replicated. Second, large participant samples will be needed to reliably detect a depletion effect. To be sure, manipulations vary in strength and dependent measures in sensitivity (which, in part, is why we used a paradigmatic approach). As seen here, descriptively, the E-task protocol showed a bigger depletion effect than the writing-task protocol.[7] Regardless, neither protocol yielded large effects.

Further, researchers may consider the role of self-reported fatigue. The current project found larger depletion effects among participants reporting more fatigue after the manipulation task, similar to earlier findings (e.g., Clarkson et al., 2010). Measuring fatigue, using manipulations that feel fatiguing, or applying manipulations known to decrease fatigue (Sripada, Kessler, & Jonides, 2014) may be worthwhile.

The current project was inspired by a previous multi-lab study of ego depletion, which reported an effect size of $d = 0.04$ (Hagger et al., 2016). A recent multi-lab test reported an effect of $d = 0.10$. That study tested the same individual differences as did we, finding little in the way of moderation (Dang et al., in press).

All told, the results from two multi-lab investigations compare similarly to the current results. The general conclusion is that the depletion effect is likely small (including zero) and not substantially moderated by theoretically-relevant dispositional differences.

Paradigmatic Replication Approach Revisited

We introduced a number of changes to the way that multi-lab replication projects typically are run, innovations aimed at increasing the knowledge gained from the project. The project used two protocols (sets of independent and dependent variables) that were not drawn from any specific study. Rather, the aim was to use permutations that befit the essence of depletion theory (that is, were paradigmatic) while allowing for

[7] By referring to protocols by their manipulation tasks, we do not mean to imply those tasks necessarily made the difference. The dependent measures may have been differentially sensitive or other factors were at work.

the possibility that the protocols may evince different outcomes and thus inform future work.

We used crowdsourcing — among topic experts and the laboratories that would be enacting them — for which manipulation and outcome tasks to use. Topic experts initially created lists of possible tasks. Subsequently, laboratories indicated whether they could execute the tasks and whether they would provide good tests of the hypothesis, which formed the basis of the tasks used.

Crowdsourcing in this way has advantages. Proponents of an effect can help identify tasks that are road-tested and reflect the theory — and ideally cut down on concerns about the methods after results are known. For replication attempts to move the field forward, it will be helpful if proponents see them as credible. Further, replications that are not direct copies of existing studies may benefit from evaluations by participating scholars to determine the tasks likely to provide good tests of the hypothesis.

Video recordings of experimenters were another novel aspect. Potential variability in execution can be a concern for multi-site projects, but also an opportunity for insights into what contributes to replication outcomes.

We followed open science practices and introduced a few of our own. The project was preregistered and data blinded for analysts' initial hypothesis tests. Outside experts provided methodological and statistical advice, another use of crowdsourcing. We put multiple layers between the project organizers and the data. Laboratories sent their data to an independent scholar who then sent them to analysts. Project organizers received the data only after initial analyses were done (Table 1).

The goals of these implementations were two-fold. One was to conduct the project in a high-quality, high-integrity manner. The other was to inspire future replication projects. If replication studies are going to be a mainstay of the field, then having more replication models can enable more suitable, relevant, and informative tests.

All studies have their limitations, and this one is no exception. We undertook a challenge by aiming to retain a large sample while introducing a new approach to replication studies to test a controversial hypothesis, the results of which were likely to have implications for the field (and for some of the authors).

One part of the project that incurred many hiccups was the preregistration, namely in terms of preregistered analyses versus analyses that were most suitable for testing the hypotheses. For instance, we did not preregister analyses at the participant level for participant-level effects (e.g., moderation by psychological states and individual differences) but should have. We could have made better preregistration choices.

The criteria for excluding participants' data also deserves mention. They were chosen with the aim of ensuring that the manipulations would elicit the intended psychological states, but we did not anticipate that they would lead to excluding nearly a third of the sample. In hindsight, perhaps we should have preregistered that we would relax some exclusion criteria if the exclusion rate exceeded a certain percentage of the total sample (e.g., 20%). More extensive pilot testing also may have helped to identify issues with the exclusion criteria prior to data collection. Additional development and validation of exclusion criteria in ego depletion research (and beyond) is sorely needed.

A last consideration is the possibility that different procedures would have yielded stronger evidence of ego depletion. Many different tasks have been used to operationalize both the independent and dependent variables in depletion studies, among which we used only four. At present, theoretical accounts generally do not indicate whether or how depletion depends on the specific manipulation or outcome tasks, but proponents of such an idea may consider high-powered, preregistered tests of that hypothesis.

Conclusion

Ego depletion is one of the most storied and, of late, questioned effects in psychological science. We embarked on a large-scale replication using two methods to manipulate self-control usage and subsequently measure it, thereby establishing the paradigmatic replication approach, a new way of testing the robustness of theoretical phenomena.

In terms of results, both the frequentist and Bayesian preregistered analyses showed no depletion effect. Exploratory Bayesian tests were inconclusive. Exploratory frequentist analyses on the full sample (without exclusions) showed a small depletion effect as well as moderation by fatigue, with a larger effect observed among depletion condition participants who reported greater fatigue. Those doubtful of the theory may see the findings as damning for the ego depletion hypothesis. Those inclined toward the theory may retort that some exploratory results suggest that there may be an effect, especially under certain conditions, although this conclusion must remain tentative.

Whether a depletion effect matters is a related but different issue. Funder and Ozer (2019) proposed that small effects (in terms of effect size) are probably more realistic than large effects, and that their value should be judged in light of the importance of the phenomenon. On that score, understanding how self-control operates seems worthy indeed.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1-48.

Baumeister, R. F., Bratslavsky, M., 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. (2016a). Strength model of self-regulation as limited resource: assessment, controversies, update. *Advances in Experimental Social Psychology, 54*, 67-127.

Baumeister, R.F., & Vohs, K. D. (2016b). Misguided effort with elusive implications. *Perspectives on Psychological Science*, *11*, 574-575. doi: 10.1177/1745691616652878.

Bullard, S. E., Fein, D., Gleeson, M. K., Tischer, N., Mapou, R. L., & Kaplan, E. (2004). The Biber cognitive estimation test. *Archives of Clinical Neuropsychology*, *19*, 835-846.

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.

Clarkson, J. J., Hirt, E. R., Jia, L., & Alexander, M. E. (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*, 28-46.

Dang, J. (2016). Commentary: A multilab preregistered replication of the ego-depletion effect. *Frontiers in Psychology, 7,* Article ID 1155.

Dang, J., Barker, P., Baumert, A., Bentvelzen M., Berkman, E. T….Zinkernagel, A. (in press). A multi-lab replication of the ego depletion effect. *Social Psychological and Personality Science.*

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

Dvorak, R. D., & Simons, J. S. (2009). Moderation of resource depletion in the self-control strength model: Differing effects of two modes of self-control. *Personality and Social Psychology Bulletin, 35,* 572-583.

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

Fein, D., Gleeson, M. K., Bullard, S., Mapou, R., & Kaplan, E. (1998, February). *The Biber Cognitive Estimation Test*. Poster presented at the annual meeting of the International Neuropsychological Society, Honolulu, HI.

Fennis, B. M., Janssen, L., & Vohs, K. D. (2009). Acts of benevolence: A limited-resource account of compliance with charitable requests. *Journal of Consumer Research, 35,* 906-924.

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.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science, 2*, 156–168. https://doi.org/10.1177/2515245919847202

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.

Gronau, Q. F., Van Erp, S., Heck, D. W., Cesario, J., Jonas, K. J., & Wagenmakers, E.-J. (2017). A Bayesian model-averaged meta-analysis of the power pose effect with informed and default priors: The case of felt power. *Comprehensive Results in Social Psychology*, *2*, 123-138.

Hagger, M. S., & Chatzisarantis, N. L. D. (2016). Commentary: "Misguided effort with elusive implications" and "sifting signal from noise with replication science". *Frontiers in Psychology, 7*, 621. https://doi.org/10.3389/fpsyg.2016.00621

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A., . . . 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.

Inzlicht, M., Gervais, W. M., & Berkman, E. T. (2015). *Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015*. Retrieved from http://ssrn.com/abstract=2659409

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.

Jeffreys, H. (1939). *Theory of probability*. Oxford: Oxford University Press.

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.

Jostmann, N. B., & Koole, S. L. (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.

Kuhl, J. (1994). *Action versus state orientation: Psychometric properties of the Action Control Scale (ACS-90).* In J. Kuhl & J. Beckmann (Eds.), Volition and Personality (pp. 47–59). Go¨ttingen, Germany: Hogrefe & Huber.

Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, *45*, 594-597.

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.

Nosek, B. A., Graham, J., Lindner, N. M., Kesebir, S., Hawkins, C. B., Hahn, C., Schmidt, K., Motyl, M., Joy-Gaba, J . A., Frazier, R., & Tenney, E. R. (2010). Cumulative and career stage impact of social-personality psychology programs and their members. *Personality and Social Psychology Bulletin, 36*, 1283-1300.

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

Scheibehenne, B., Gronau, Q., Jamil, T., & Wagenmakers, E.-J. (2017). Fixed or random? A resolution through model-averaging: Reply to Carlsson, Schimmack, Williams, and Burkner. *Psychological Science, 28,* 1698-1701*.*

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

Schmeichel, B. J., Vohs, K. D., & Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, *85*, 33-46.

Spellman, B. A., & Kahneman, D. (2018). What the replication reformation wrought: Comment on Zwaan et al. *Behavioral and Brain Sciences, 41*, E149. https://doi.org/10.1017/s0140525x18000857

Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science*, *25*, 1227-1234.

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 Erp, S., Verhagen, J., Grasman, R.P.P.P. and Wagenmakers, E.-J., 2017. Estimates of between-study heterogeneity for 705 meta-analyses reported in

Psychological Bulletin from 1990–2013. *Journal of Open Psychology Data*, *5*, p. 4. DOI:http://doi.org/10.5334/jopd.33

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1-48.

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

Table 1. *Paradigmatic Replication Approach: Goals, Strategies, and Rationales*

Goals	Strategies	Rationale(s)
Formulation stage		
Identify representative tasks	Crowdsource with area experts • Create list of possible IV and DV tasks deemed paradigmatic for testing the hypothesis	Collect diversity of possible methods Get help from experienced researchers in topic area
Select sound methods	Prioritize the operationalization of psychological states, not whether a specific study replicates Tasks need not mimic a published study	Not tied to other scholars' choices and methods Can adjust for project goals, labs, participant characteristics
Boost commitment from participating laboratories	Crowdsource with participating labs • Assess whether tasks are deemed to be executable and effective • Gather preferences for possible tasks	Winnow down the set of possible tasks with scholars who will be executing the study Enable scholars who will be executing the study to have some say in the methods
Ensure rigorous design and analysis choices	Assemble methods and statistics Advisory Board • Understand implications of methodological and statistical options before preregistration • Perform main hypothesis-testing analyses • Consider using both frequentist and Bayesian approaches	Open science practices Expand skill set beyond what project leaders bring Increase information value of results
Study preparation stage		
Public statement(s) of intent	Preregistration of hypotheses, methods, participant exclusion criteria, and specify conclusions given different possible results	Open science practices Reduce researcher degrees of freedom
Methods testing and practice	Video recordings of how to conduct the study Write and review scripts for experimenters to follow	Reduce variation in procedural execution
Team building	Virtual meetings with all members of participating labs	Address questions, reinforce procedural details, and bridge gap between project leaders and data collection labs

Post-data collection stage		
Ensure data integrity	Labs send data to independent handler. Data handler: • Merges data files • Blinds outcome measures • Sends master dataset to Advisory Board	Project managers do not receive data until initial analyses are done Ensure data integrity and increase confidence in the results
Increase information value of data	After designated data analysts conduct confirmatory tests, lead authors can suggest exploratory analyses	Follow up on relevant hypothesis tests Perform tests that were unanticipated or underspecified in preregistration

Note: IV stands for independent variable. DV stands for dependent variable.

Table 2. *Exclusion Counts for Each Preregistered Criteria by Protocol*

Criteria	E-Task Protocol	Writing Task Protocol
Errors on last completed E-task paragraph (Page 1)	159	NA
Errors on last completed E-task paragraph (Page 2)	133	NA
Knew puzzles were unsolvable	42	NA
Used few words in story	NA	7
Used forbidden letters in story	NA	83
Invalid responses on CET	NA	0
Non-native speakers	95	223
First three participants	111	96
Used phone during study	79	63
Belligerent	2	3
Distressed/distraught	9	7
Disruption or other unanticipated deviation	19	11
Other exclusions	174	34
TOTALS	823	527

Note: NA stands for not applicable. In total *n* = 1068 were excluded in accordance with preregistered exclusion criteria. Some participants (*n* = 237) failed multiple exclusion criteria. See SOM for additional details.

Table 3. *Manipulation Checks: Descriptive Statistics and Frequentist Meta-Analytic Tests of Experimental Conditions*

	M (SD)						
Variable	**Depletion**	**Non-depletion**	**FE Average**	**CI**	**RE Average**	**CI**	**I^2**
Effort index	4.52 (1.74)	2.59 (1.11)	1.31***	[1.22,1.40]	1.64***	[1.18, 2.09]	95.65%
Frustration	3.81 (2.01)	2.04 (1.39)	0.99***	[0.90, 1.08]	1.14***	[0.77, 1.50]	93.95%
Fatigue index	3.29 (1.53)	2.89 (1.53)	0.25***	[0.17, 0.33]	0.24**	[0.07, 0.41]	76.60%
Motivation index	5.25 (1.20)	5.14 (1.27)	0.05	[-0.03, 0.13]	0.04	[-0.06, 0.13]	30.33%

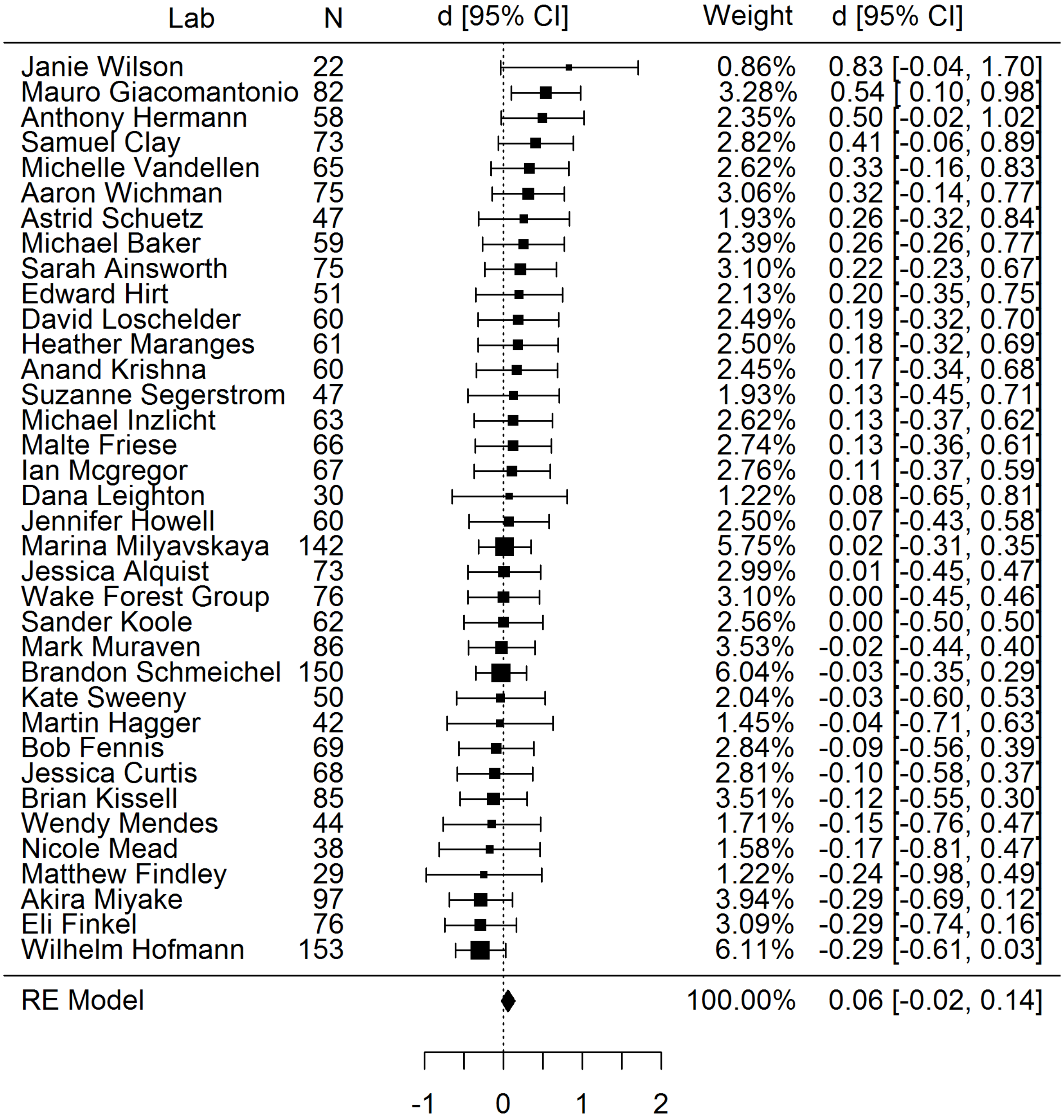
Note: *N* = 2463 (*k* = 36), with the exception that frustration ratings were missing for two participants. Sample size varies from total sample size due to missing data. Condition coded such that 0 = non-depletion, 1 = depletion condition. Higher numbers indicate that participants in the depletion condition reported stronger feelings than participants in the non-depletion condition. All tests were confirmatory (preregistered). Means and *SD* are from unstandardized scales; range 1 (*not at all*) to 7 (*very*). FE indicates fixed-effects models; RE indicates random-effects models. *CI* indicates 95% confidence intervals. ** *p* < .01; *** *p* < .001

Table 4. *Depletion Effect: Frequentist Meta-Analyses*

		Random-effects meta-analysis			Fixed-effects meta-analysis	
DV	***N***	**d**	**CI**	**I^2%**	**d**	**CI**
Overall depletion effect	2461	0.06	[-0.02, 0.14]	2.54	0.06	[-0.02, 0.14]
Overall figure tracing performance	1216	0.12	[-0.01, 0.24]	15.16	0.11	[-0.00, 0.22]
Figure tracing duration	1216	0.15 *	[0.02, 0.29]	28.46	0.13 *	[0.02, 0.25]
Figure tracing attempts	1217	0.05	[-0.06, 0.17]	0	0.05	[-0.06, 0.17]
Cognitive Estimation Test	1245	0.01	[-0.10, 0.12]	0	0.01	[-0.10, 0.12]

Note: Sample sizes vary due to missing data. For overall depletion effect analyses, *k* = 36; figure tracing analyses, *k* = 20; Cognitive Estimation Test analyses, *k* = 16. Condition coded such that 0 = non-depletion, 1 = depletion condition. Higher numbers indicate evidence of a depletion effect (i.e., self-control was worse in the depletion condition). DV stands for dependent variable. FE indicates fixed-effects models; RE indicates random-effects models. CI indicates 95% confidence intervals. * *p* < .05

Figure 1. *Forest Plot of Performance Outcome by Laboratory.* The box plots and numerical values illustrate the same effect size estimates. For the plots, the size of the box represents its weighted contribution to the overall effect and its whiskers display 95% CIs. The dotted line represents a zero effect size. Numerical values show standardized mean differences between depletion and non-depletion conditions expressed in Cohen’s *d* (with 95% CIs). The diamond is the overall meta-analytic effect derived from a random-effects model.



Lab	N	Weight	d [95% CI]
Janie Wilson	22	0.86%	0.83 [-0.04, 1.70]
Mauro Giacomantonio	82	3.28%	0.54 [0.10, 0.98]
Anthony Hermann	58	2.35%	0.50 [-0.02, 1.02]
Samuel Clay	73	2.82%	0.41 [-0.06, 0.89]
Michelle Vandellen	65	2.62%	0.33 [-0.16, 0.83]
Aaron Wichman	75	3.06%	0.32 [-0.14, 0.77]
Astrid Schuetz	47	1.93%	0.26 [-0.32, 0.84]
Michael Baker	59	2.39%	0.26 [-0.26, 0.77]
Sarah Ainsworth	75	3.10%	0.22 [-0.23, 0.67]
Edward Hirt	51	2.13%	0.20 [-0.35, 0.75]
David Loschelder	60	2.49%	0.19 [-0.32, 0.70]
Heather Maranges	61	2.50%	0.18 [-0.32, 0.69]
Anand Krishna	60	2.45%	0.17 [-0.34, 0.68]
Suzanne Segerstrom	47	1.93%	0.13 [-0.45, 0.71]
Michael Inzlicht	63	2.62%	0.13 [-0.37, 0.62]
Malte Friese	66	2.74%	0.13 [-0.36, 0.61]
Ian Mcgregor	67	2.76%	0.11 [-0.37, 0.59]
Dana Leighton	30	1.22%	0.08 [-0.65, 0.81]
Jennifer Howell	60	2.50%	0.07 [-0.43, 0.58]
Marina Milyavskaya	142	5.75%	0.02 [-0.31, 0.35]
Jessica Alquist	73	2.99%	0.01 [-0.45, 0.47]
Wake Forest Group	76	3.10%	0.00 [-0.45, 0.46]
Sander Koole	62	2.56%	0.00 [-0.50, 0.50]
Mark Muraven	86	3.53%	-0.02 [-0.44, 0.40]
Brandon Schmeichel	150	6.04%	-0.03 [-0.35, 0.29]
Kate Sweeny	50	2.04%	-0.03 [-0.60, 0.53]
Martin Hagger	42	1.45%	-0.04 [-0.71, 0.63]
Bob Fennis	69	2.84%	-0.09 [-0.56, 0.39]
Jessica Curtis	68	2.81%	-0.10 [-0.58, 0.37]
Brian Kissell	85	3.51%	-0.12 [-0.55, 0.30]
Wendy Mendes	44	1.71%	-0.15 [-0.76, 0.47]
Nicole Mead	38	1.58%	-0.17 [-0.81, 0.47]
Matthew Findley	29	1.22%	-0.24 [-0.98, 0.49]
Akira Miyake	97	3.94%	-0.29 [-0.69, 0.12]
Eli Finkel	76	3.09%	-0.29 [-0.74, 0.16]
Wilhelm Hofmann	153	6.11%	-0.29 [-0.61, 0.03]
RE Model		100.00%	0.06 [-0.02, 0.14]

Figure 2. *Bayesian Forest Plot of Performance Outcome by Laboratory*. The values listed under BF_{+0} indicate relative support for the depletion hypothesis versus a hypothesis that there is no effect. Diamonds indicate overall effect sizes from meta-analytic models using fixed-effects, random-effects, and one that combined both approaches.

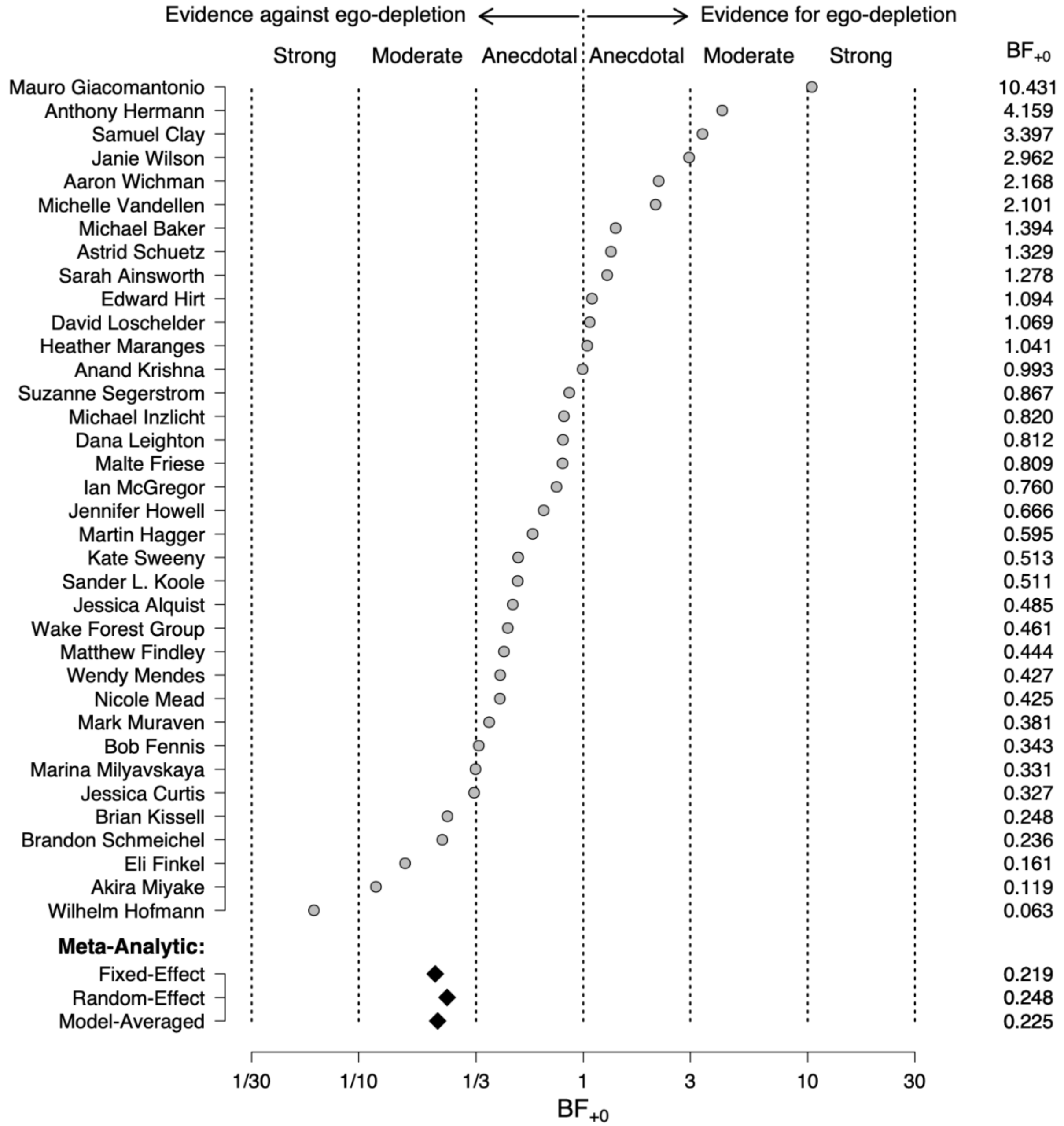
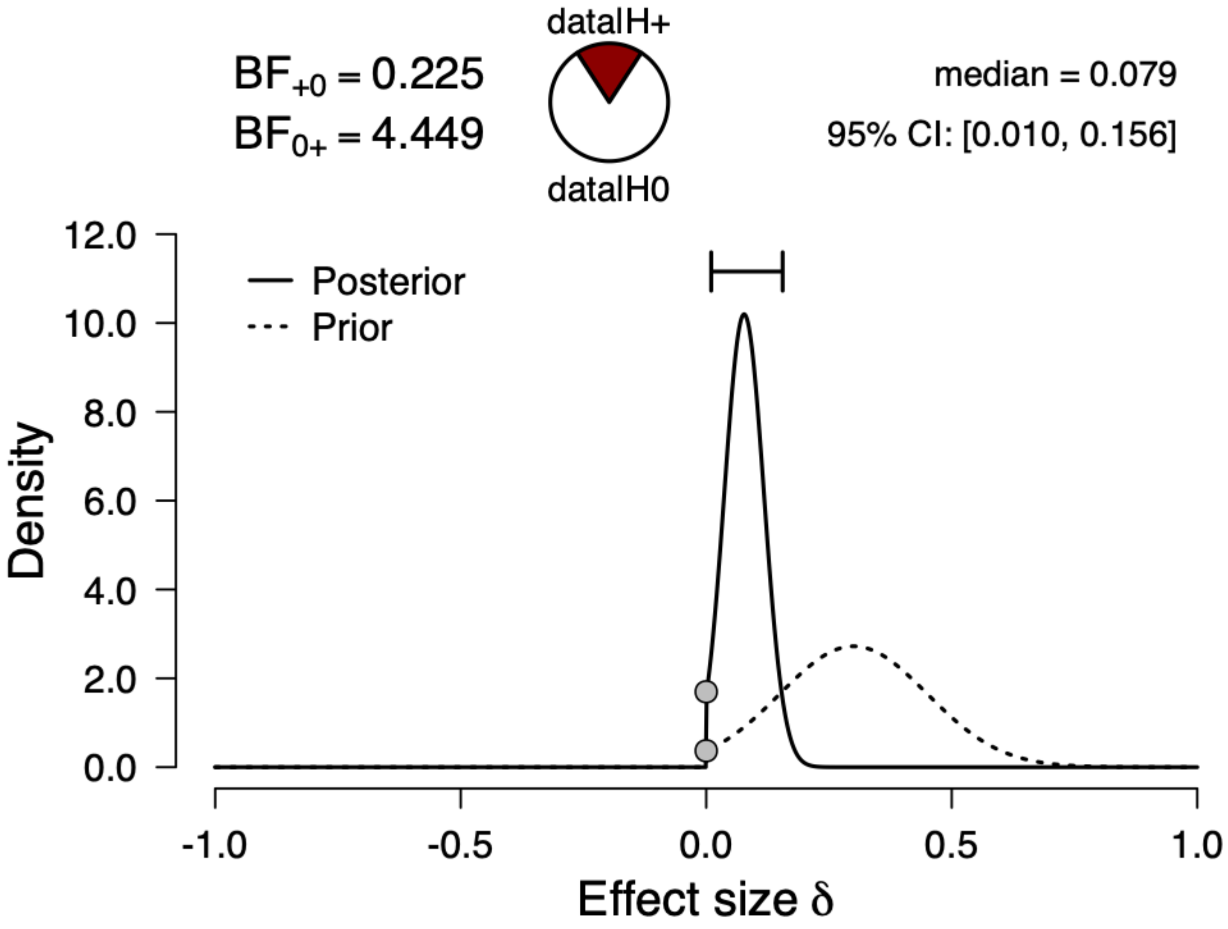


Figure 3. *Tests of the Model-Averaged Meta-Analytic Effect Size Posterior and Bayes Factor*. The dotted line indicates the informed prior effect size distribution and the solid line indicates the model-averaged meta-analytic posterior effect size distribution. Roughly-speaking, the peak of the shape indicates the likelihood of the effect size and its width indicates variance.



A Multi-Site Preregistered, Paradigmatic Test of the Ego Depletion Effect

SUPPLEMENTARY ONLINE MATERIALS

Contents

Exploratory Analyses

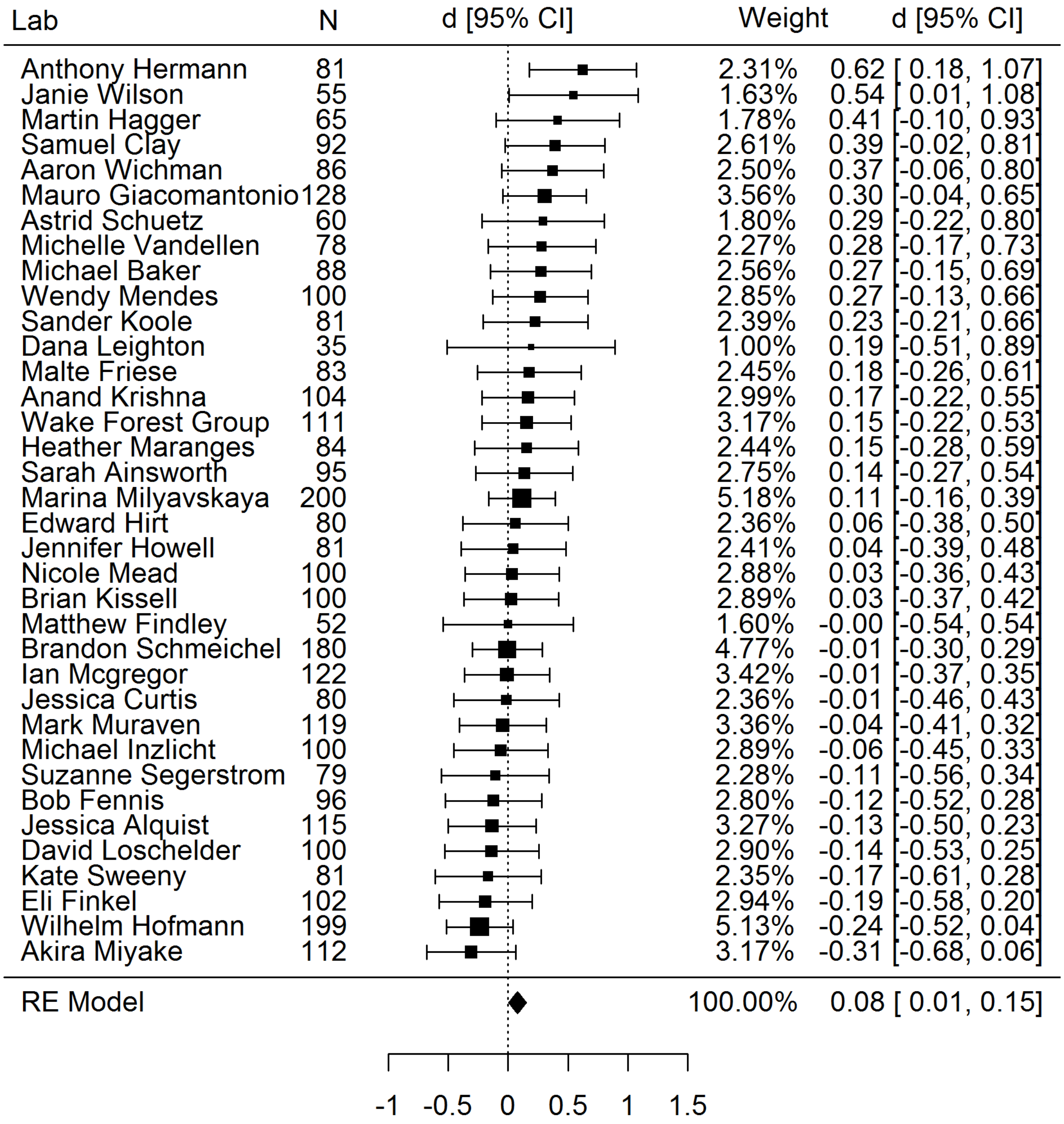
Depletion effect

Frequentist analyses. Analyses based on the full dataset were not preregistered, but the rate of exclusions far exceeded expectations. We therefore decide to conduct exploratory analyses using the full dataset.

Meta-analyses of the full dataset revealed a small significant effect in line with predictions (RE: d = 0.08, 95% CI [0.01, 0.15]; FE: d = 0.07, 95% CI [0.01, 0.14]; I^2 = 11.69%; Figure S1). This effect was observed for both random- and fixed-effects models. Experimental protocol did not appear to moderate the depletion effect, RE: intercept d = 0.08 [0.00, 0.15], moderator b = -0.07 95% CI [-0.22, 0.07], I^2 = 13.90%.

We also tested whether there was evidence of an overall depletion effect using multilevel regression approaches that nested the individual-level data within laboratories in random-intercept mixed models. In the reduced sample (excluding 1068 participants, following preregistered rules), task performance did not differ by depletion condition, b = 0.09 CI [-0.01, 0.19]. In the full sample (when participants marked for exclusion were included), the effect of depletion condition was statistically significant but small (Table S1).

Figure S1. *Forest Plot of Performance Outcome by Laboratory: Full Sample.* The box plots and numerical values illustrate the same effect size estimates. For the plots, the size of the box represents its weighted contribution to the overall effect and its whiskers display 95% CIs. The dotted line represents a zero effect size. Numerical values show standardized mean differences between depletion and non-depletion conditions expressed in Cohen's *d* (with 95% CIs). The diamond is the overall meta-analytic effect derived from a random-effects model.



Lab	N	Weight	d [95% CI]
Anthony Hermann	81	2.31%	0.62 [0.18, 1.07]
Janie Wilson	55	1.63%	0.54 [0.01, 1.08]
Martin Hagger	65	1.78%	0.41 [-0.10, 0.93]
Samuel Clay	92	2.61%	0.39 [-0.02, 0.81]
Aaron Wichman	86	2.50%	0.37 [-0.06, 0.80]
Mauro Giacomantonio	128	3.56%	0.30 [-0.04, 0.65]
Astrid Schuetz	60	1.80%	0.29 [-0.22, 0.80]
Michelle Vandellen	78	2.27%	0.28 [-0.17, 0.73]
Michael Baker	88	2.56%	0.27 [-0.15, 0.69]
Wendy Mendes	100	2.85%	0.27 [-0.13, 0.66]
Sander Koole	81	2.39%	0.23 [-0.21, 0.66]
Dana Leighton	35	1.00%	0.19 [-0.51, 0.89]
Malte Friese	83	2.45%	0.18 [-0.26, 0.61]
Anand Krishna	104	2.99%	0.17 [-0.22, 0.55]
Wake Forest Group	111	3.17%	0.15 [-0.22, 0.53]
Heather Maranges	84	2.44%	0.15 [-0.28, 0.59]
Sarah Ainsworth	95	2.75%	0.14 [-0.27, 0.54]
Marina Milyavskaya	200	5.18%	0.11 [-0.16, 0.39]
Edward Hirt	80	2.36%	0.06 [-0.38, 0.50]
Jennifer Howell	81	2.41%	0.04 [-0.39, 0.48]
Nicole Mead	100	2.88%	0.03 [-0.36, 0.43]
Brian Kissell	100	2.89%	0.03 [-0.37, 0.42]
Matthew Findley	52	1.60%	-0.00 [-0.54, 0.54]
Brandon Schmeichel	180	4.77%	-0.01 [-0.30, 0.29]
Ian Mcgregor	122	3.42%	-0.01 [-0.37, 0.35]
Jessica Curtis	80	2.36%	-0.01 [-0.46, 0.43]
Mark Muraven	119	3.36%	-0.04 [-0.41, 0.32]
Michael Inzlicht	100	2.89%	-0.06 [-0.45, 0.33]
Suzanne Segerstrom	79	2.28%	-0.11 [-0.56, 0.34]
Bob Fennis	96	2.80%	-0.12 [-0.52, 0.28]
Jessica Alquist	115	3.27%	-0.13 [-0.50, 0.23]
David Loschelder	100	2.90%	-0.14 [-0.53, 0.25]
Kate Sweeny	81	2.35%	-0.17 [-0.61, 0.28]
Eli Finkel	102	2.94%	-0.19 [-0.58, 0.20]
Wilhelm Hofmann	199	5.13%	-0.24 [-0.52, 0.04]
Akira Miyake	112	3.17%	-0.31 [-0.68, 0.06]
RE Model		100.00%	0.08 [0.01, 0.15]

Table S1. *Depletion Effect: Exploratory Frequentist Meta-Analyses and Multi-Level Models*

		Random-effects meta-analysis			Fixed-effects meta-analysis		Multi-level regression	
DV	***N***	***d***	**CI**	**I^2 %**	***d***	**CI**	***b***	**CI**
Overall depletion effect	3524	0.08 *	[0.01, 0.15]	11.69	0.07 *	[0.01, 0.14]	0.11 *	[0.02, 0.20]
Overall figure tracing performance	1847	0.12 *	[0.01, 0.23]	27.23	0.10 *	[0.01, 0.20]	0.18 *	[0.03, 0.32]
Figure tracing duration	1847	0.14 *	[0.01, 0.27]	46.83	0.12 *	[0.03, 0.21]	0.11 *	[0.02, 0.20]
Figure tracing attempts	1848	0.06	[-0.04, 0.15]	0	0.06	[-0.04, 0.15]	0.07	[-0.02, 0.15]
Cognitive Estimation Test	1677	0.04	[-0.06, 0.13]	0	0.04	[-0.06, 0.13]	0.04	[-0.06, 0.13]

Note: Results pertain to the entire sample. Sample sizes vary due to missing data. For overall depletion effect analyses, *k* = 36; figure tracing analyses, *k* = 20; Cognitive Estimation Test analyses, *k* = 16. Condition coded such that 0 = non-depletion, 1 = depletion condition. Higher numbers indicate evidence of a depletion effect (i.e., self-control was worse in the depletion condition). DV stands for dependent variable. FE indicates fixed-effects models; RE indicates random-effects models. *CI* indicates 95% confidence intervals. Multi-level models nested participants' data within labs and used a random intercept for labs. * *p* < .05

Bayesian analyses. We next turn to the model-averaged meta-analytic Bayes factor (which corresponds closely to the fixed- and random-effects Bayes factors; Figure S2). The results indicated that the data are 1.33 times more likely under the point-null hypothesis (which states that the effect is absent) than under the one-sided informed alternative hypothesis (which states that the effect is present), suggesting that two models predict the data almost equally well. Although the full sample data provided no basis for shifting beliefs towards or away from either hypothesis, the posterior distribution addressed the magnitude of the effect if it is present.

To take into account the findings from all laboratories simultaneously, we considered the results of the model-averaged meta-analysis. Figure S3 displays the model-averaged meta-analytic posterior for effect size as a solid line; the dotted line indicates the informed prior distribution. As shown, the data have shifted our beliefs about the effect size of ego depletion toward zero. Specifically, the posterior median was 0.087 with a central 95% credible interval ranging from 0.023 to 0.152 (Figure S3).

Figure S2. *Bayesian Forest Plot of Performance Outcome by Laboratory: Full Sample.* The values listed under BF+0 indicate relative support for the depletion hypothesis versus a hypothesis that there is no effect. Diamonds indicate overall effect sizes from meta-analytic models using fixed-effects, random-effects, and one that combined both approaches.

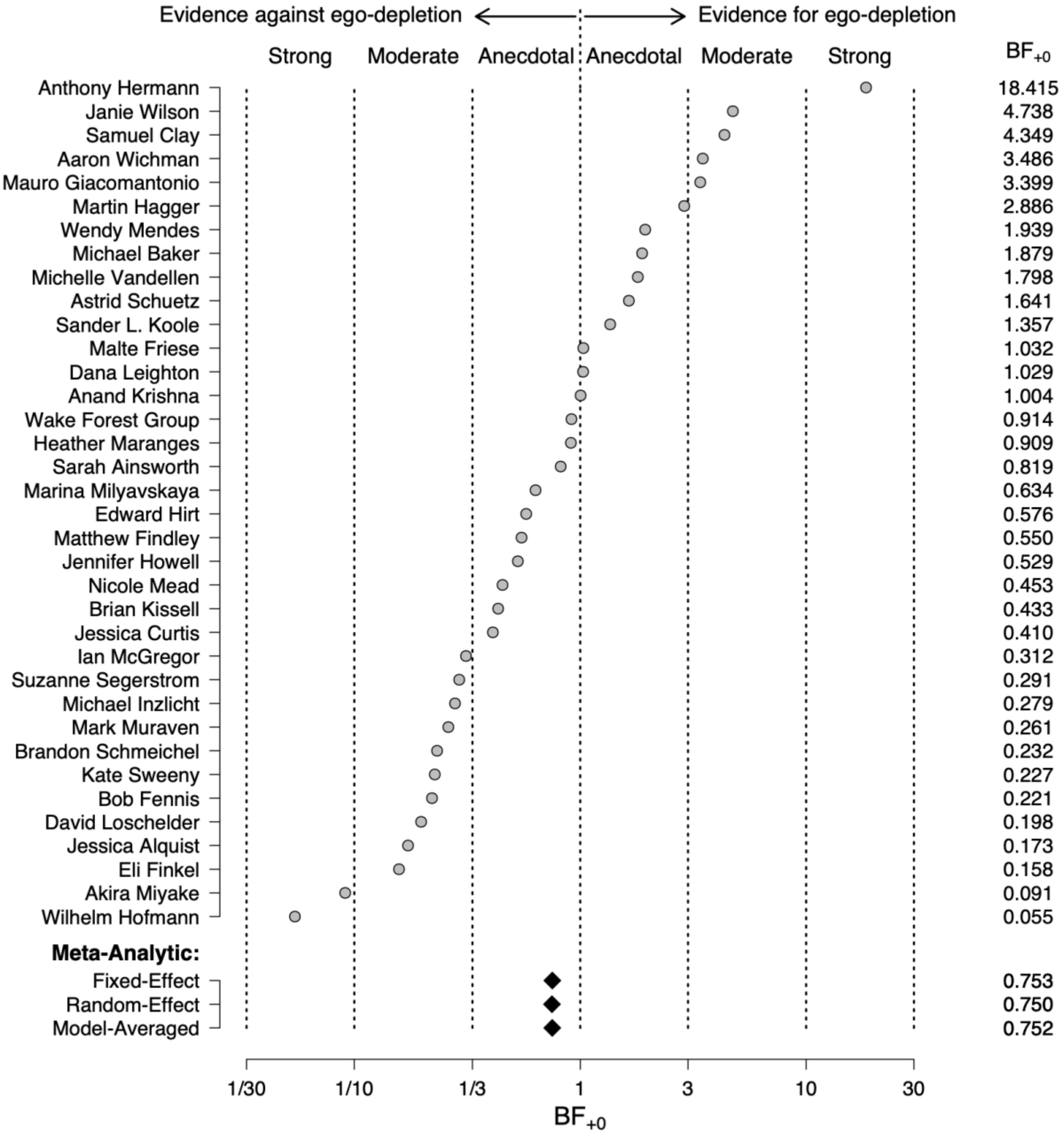
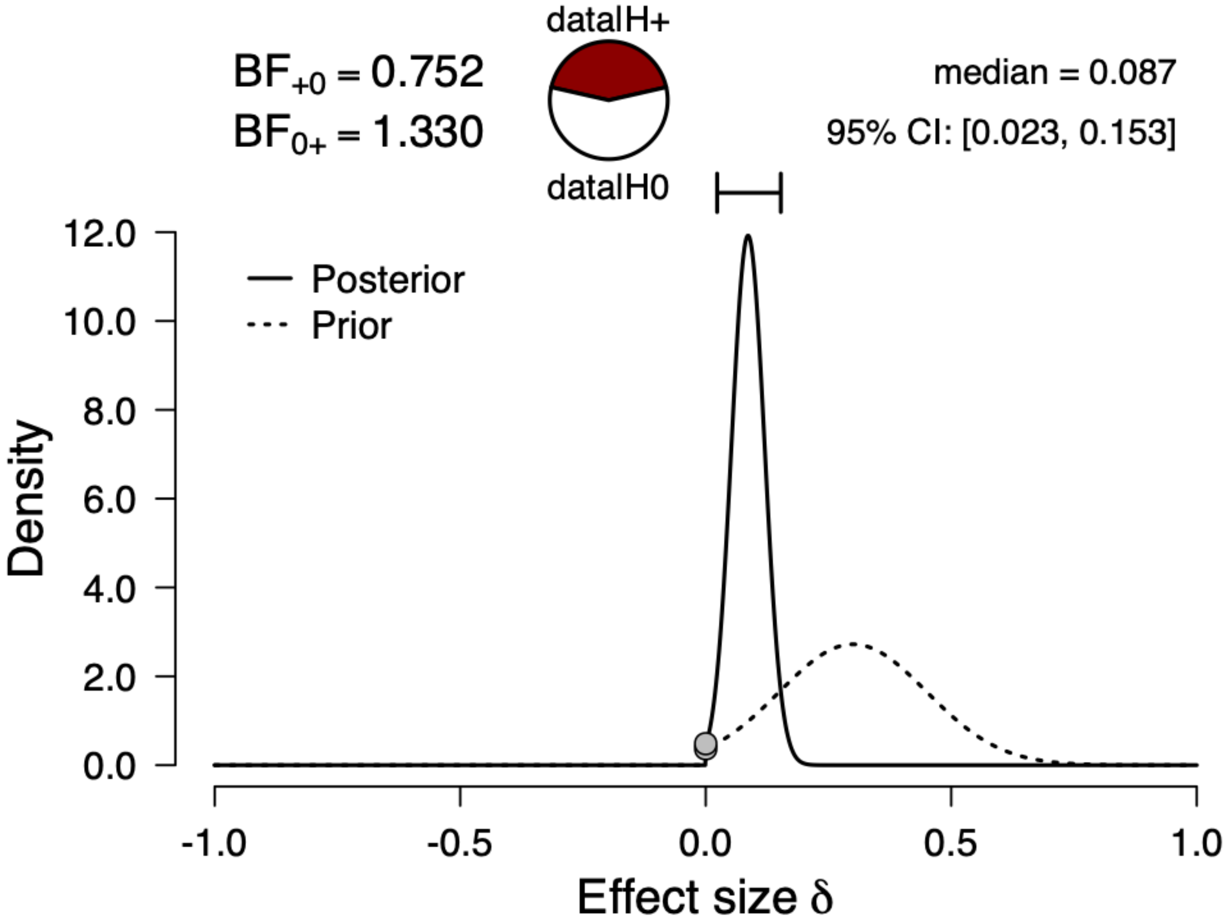


Figure S3. *Exploratory Tests of Model-Averaged Meta-Analytic Effect Size Posterior and Bayes Factor: Full Sample*. The dotted line indicates the informed prior effect size distribution and the solid line indicates the model-averaged meta-analytic posterior effect size distribution. Roughly-speaking, the peak of the shape indicates the likelihood of the effect size and its width indicates variance.



Moderators of the Depletion Effect

Protocol type. The main article reports confirmatory meta-analytical tests on the reduced sample (after preregistered exclusions; see Table 4). Here, we supplant those with parallel, exploratory results on the full sample and multi-level regressions on both samples.

Full sample: We examined the two components of the figure tracing task separately, the number of sheets participants used (as an indicator of attempts) and time spent working on the task (in seconds). Examining the two components separately, the effect of depletion condition on number of attempts was not statistically significant (Table S1; unstandardized descriptives: non-depletion condition *M* = 19.71, *SD* = 10.05; depletion condition *M* = 19.09, *SD* = 10.21).

There was a significant effect of depletion condition on duration in the full sample (unstandardized descriptives: non-depletion condition *M* = 988.87s, *SD* = 283.95; depletion condition *M* = 960.03s, *SD* = 298.52). These exploratory analyses showed that participants in the depletion condition gave up on the figure tracing task around 28s sooner than participants in the non-depletion condition (Table S1).

For the combined measure of figure tracing duration and attempts in the E-task protocol, there was a statistically significant effect in the full sample, as judged by both the meta-analytic and multi-level regression approaches. Participants in the depletion condition had lower figure tracing scores than did participants in the non-depletion condition (Table S1).

We conducted meta-analytic and multi-level analyses within the writing task protocol, which used the Cognitive Estimation Test (CET) as the performance measure.

The results were non-significant (unstandardized descriptives: non-depletion condition *M* = 1.32, *SD* = 0.23; depletion condition *M* = 1.31, *SD* = 0.24; Table S1).

Reduced sample. Multi-level regression models analyzed the reduced sample's performance within each protocol. For overall figure tracing scores, the effect of condition was not significant, b = 0.17, 95% CI [-0.01, 0.34].

As mentioned, that score has two elements. Breaking them down, the effect of condition on number of attempts was not statistically significant (*b* = 0.06, 95% CI [-.05, 0.17]; unstandardized descriptives: non-depletion condition *M* = 19.87, *SD* = 9.92; depletion condition *M* = 19.36, *SD* = 10.41).

As in the full sample, the effect of depletion condition on duration was significant in the reduced sample (*b* = 0.11, 95% CI [.01, 0.21]; unstandardized descriptives: non-depletion condition *M* = 1012.20s, *SD* 266.30; depletion condition *M* = 985.10s, *SD* = 283.52).

A last set of exploratory analyses regarded the depletion manipulation's effect on CET performance. As in the full sample, the effect of depletion condition was non-significant in the reduced sample (*b* = 0.01, 95% CI [-.09, 0.12]; unstandardized descriptives: non-depletion condition *M* = 1.34, *SD* = .23; depletion condition *M* = 1.34, *SD* = .23).

States and traits. We also examined whether self-reported states captured by the manipulation check items (e.g., fatigue, effort) and individual difference measures (i.e., trait self-control; willpower beliefs; action orientation) acted as moderators of the depletion effect. Because self-reported traits and states are best modeled as individual-level data, multilevel regressions were used as opposed to meta-analytic analyses (Table S2).

The only significant moderator was the fatigue index, which was evident in both the reduced and full samples. The depletion effect was larger for participants who reported being more fatigued by the manipulation task.

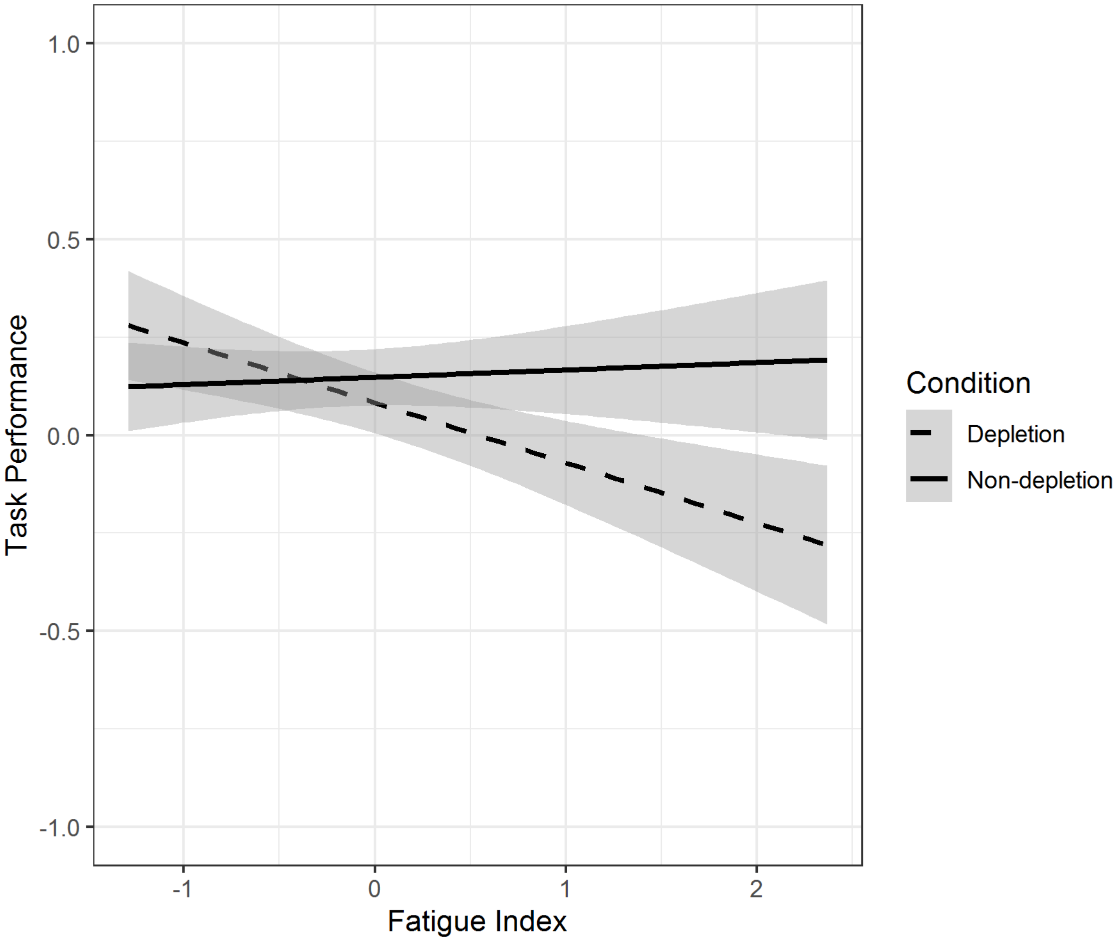
For the reduced sample (after exclusions), simple-slope analyses revealed that within the range of the data, the depletion effect was significant in a region from a standardized score of 0.15 on the fatigue index to the sample maximum of 2.37 (Figure S4). The magnitude of the depletion effect was $b = 0.23$, $SE = 0.07$, $p = .001$, at the 75th percentile of fatigue (0.84). For the full sample, the magnitude of the depletion effect at the 75th percentile of fatigue (0.84) was $b = 0.21$, $SE = 0.06$, $p < .001$.

Table S2. *Potential Moderators of the Depletion Effect: Frequentist Multi-Level Models*

				Intercept		**Depletion manipulation**		**Moderator**		**Interaction**	
Moderator variable	**Moderator type**	**Sample**	***N***	***b***	**CI**	***b***	**CI**	***b***	**CI**	***b***	**CI**
Protocol^{a}	Study design	Reduced	2461	0.08	[-0.07, 0.23]	0.09	[-0.01, 0.19]	0.02	[-0.28, 0.32]	-0.16	[-0.36, 0.05]
		Full	3524	-0.04	[-0.19, 0.10]	0.11*	[0.02, 0.20]	0.06	[-0.22, 0.34]	-0.14	[-0.32, 0.04]
Effort index	Manipulation check	Reduced	2461	0.09	[-0.07, 0.24]	0.03	[-0.09, 0.16]	-0.02	[-0.11, 0.07]	-0.07	[-0.22, 0.09]
		Full	3523	-0.03	[-0.17, 0.11]	0.04	[-0.07, 0.15]	-0.03	[-0.11, 0.05]	-0.07	[-0.21, 0.06]
Fatigue index	Manipulation check	Reduced	2461	0.10	[-0.05, 0.24]	0.08	[-0.02, 0.18]	-0.15***	[-0.23, -0.07]	0.18**	[0.07, 0.29]
		Full	3523	-0.03	[-0.16, 0.11]	0.09	[-0.00, 0.18]	-0.15***	[-0.21, -0.08]	0.14**	[0.05, 0.24]
Frustration	Manipulation check	Reduced	2459	0.12	[-0.03, 0.27]	0.02	[-0.13, 0.10]	-0.11**	[-0.18, -0.03]	0.06	[-0.07, 0.19]
		Full	3521	-0.00	[-0.14, 0.14]	0.03	[-0.07, 0.13]	-0.10**	[-0.17, -0.04]	0.03	[-0.08, 0.14]
Action Orientation	Individual difference	Reduced	2356	0.08	[-0.07, 0.24]	0.08	[-0.02, 0.19]	-0.12	[-0.43, 0.20]	-0.07	[-0.51, 0.37]
		Full	3395	-0.04	[-0.18, 0.11]	0.11*	[0.02, 0.20]	-0.17	[-0.44, 0.10]	-0.03	[-0.42, 0.35]
Implicit Willpower Theory	Individual difference	Reduced	2341	0.05	[-0.10, 0.20]	0.09	[-0.01, 0.19]	0.01	[-0.07, 0.10]	0.02	[-0.09, 0.14]
		Full	3315	-0.05	[-0.19, 0.10]	0.09	[-0.00, 0.18]	0.02	[-0.06, 0.09]	0.04	[-0.06, 0.14]
Trait Self-Control	Individual difference	Reduced	2444	0.07	[-0.08, 0.22]	0.10	[-0.00, 0.20]	0.00	[-0.12, 0.12]	0.01	[-0.15, 0.17]
		Full	3490	-0.05	[-0.19, 0.09]	0.12**	[0.03, 0.21]	-0.01	[-0.11, 0.10]	0.03	[-0.11, 0.17]

Note: These tests are exploratory. Sample sizes vary due to missing data. Condition coded such that 0 = non-depletion, 1 = depletion condition. Results are raw beta weights (*b*) from random-effects multi-level mixed models; CI indicates 95% confidence intervals. Participants' data were nested within lab with random intercepts for labs and separate regression models were used for each moderator. Individual differences scores were mean-centered. [a] Contrast-coded, -.5 = E-task, .5 = Writing task. * $p < .05$

Figure S4. *Exploratory Test of Moderation of Task Performance by Depletion Condition and Self-Reported Fatigue: Reduced Sample*. The figure represents the interaction of depletion condition x fatigue scores on task performance with 95% confidence bands. Task performance was standardized and ranged from -5.54 to 7.05 (only the region from -1 to 1 is displayed). The fatigue index is an average of standardized ratings of fatigue and tiredness.



Secondary moderator analyses. We tested whether the depletion effect was moderated by: the number of depletion studies published by the principal investigator (PI) through 2016 (as counted independently by KV and BS), the number of total publications by the PI through 2016 (as counted by KV and BS), experimenter behavior (as rated by two independent coders of the videos submitted by each laboratory), and laboratory location (North American countries versus other countries). The latter moderator was chosen because many published depletion studies were conducted in North America so it was plausible that location might make a difference in the outcome. The only significant moderator in these analyses was the role of experimenter behavior in the full sample (Table S3). Experimenter behavior was not a significant moderator in the meta-analytic results in the full sample or in the meta-analytic or multi-level regression results in the reduced sample.

Exploratory multi-level regression analyses using the full sample showed an additional interaction of depletion condition and codings of experimenter behavior on task performance, *b* = -0.25, 95% CI [-0.45, -0.05]. The main effect of condition was significant in this model, *b* = 0.11, 95% CI [0.02, 0.20], and so was the main effect of experimenter behavior scores, *b* = 0.22, 95% CI [0.00, 0.43]. Simple slopes analyses of the interaction showed that experimenter behavior had no effect on performance in the non-depletion condition, *b* = -0.03, *SE* = 0.11, *p* = .776, but in the depletion condition, performance was worse when experimenters' behavior was rated lower, *b* = 0.22, *SE* = 0.11, *p* = .046. For experimenter behavior scores at the sample median (0.16) or above (that is, experimenters who were at least moderately professional, at ease, and stuck to the script), there was no depletion effect, *b* = 0.06, *SE* = 0.05, *p* = .183. Below-average experimenter behavior scores were however related to the magnitude of the depletion effect, *b* = 0.18, *SE* = 0.06, *p* = .001, at the 25th percentile of experimenter behavior scores (-0.29).

We preregistered our intention to test moderation of the depletion effect by experimenters' awareness of the depletion hypothesis or whether investigators had a Ph.D. We did not conduct these analyses because we did not solicit experimenters' knowledge of the depletion hypothesis prior to some laboratories initiating data collection and because there was very little variance in highest degree obtained. We also preregistered that we would test whether exclusion of participants based on the dependent measure differed as a function of depletion condition. The test, however, turned out to be inapplicable because the exclusion criteria were not set up to enable it.

Table S3. *Potential Depletion Effect Moderators: Exploratory Frequentist Multi-Level Models*

			Intercept		**Depletion manipulation**		**Moderator**		**Interaction**	
Moderator variable	**Sample**	***N***	***b***	**CI**	***b***	**CI**	***b***	**CI**	***b***	**CI**
Experimenter behavior	Reduced	2396	0.08	[-0.07, 0.23]	0.09	[-0.01, 0.19]	0.24*	[0.00, 0.48]	-0.19	[-0.42, 0.04]
	Full	3441	-0.04	[-0.18, 0.10]	0.11*	[0.02, 0.20]	0.22*	[0.00, 0.43]	-0.25*	[-0.45, -0.05]
Depletion studies count	Reduced	2461	0.08	[-0.09, 0.25]	0.13*	[0.01, 0.24]	-0.00	[-0.02, 0.01]	-0.01	[-0.01, 0.00]
	Full	3524	-0.07	[-0.23, 0.09]	0.15*	[0.05, 0.25]	0.00	[-0.01, 0.02]	-0.01	[-0.01, 0.00]
Publication count	Reduced	2461	0.08	[-0.07, 0.22]	0.09	[-0.01, 0.19]	0.00	[-0.00, 0.01]	-0.00*	[-0.01, -0.00]
	Full	3524	-0.05	[-0.19, 0.09]	0.11*	[0.02, 0.20]	0.00	[-0.00, 0.00]	-0.00	[-0.00, 0.00]
Lab location[a]	Reduced	2461	0.12	[-0.16, 0.40]	0.06	[-0.14, 0.25]	-0.06	[-0.39, 0.27]	0.05	[-0.18, 0.27]
	Full	3524	-0.05	[-0.31, 0.22]	0.08	[-0.08, 0.25]	0.00	[-0.31, 0.31]	0.04	[-0.16, 0.24]

Note: These tests are exploratory. All moderators are at the level of the lab. Sample sizes depart slightly from total sample sizes due to missing data. Condition coded such that 0 = non-depletion, 1 = depletion condition. Participants' data were nested within lab with random intercepts for labs and separate regression models were used for each moderator. Experimenter behavior scores were mean-centered. Depletion studies count was the number of published depletion studies by each Primary Investigator. Results are raw beta weights (*b*) from random-effects multi-level mixed models; CI indicates 95% confidence intervals. [a] Dummy-coded, 0 = Outside North America, 1 = North America. * *p*<.05

Full sample manipulation checks

We conducted the same meta-analytic tests reported in the main text on the full sample of participants (i.e., no exclusions). Using the index of effort and difficulty ratings, the manipulation worked as intended (Table S4). We tested whether effort ratings differed by protocol, coded such that the intercept (*d* = 1.69, 95% [1.59, 1.79], I^2 = 36.09%) represents the average effect across both protocols (-.5 = E-task; .5 = Writing task). As in the confirmatory reduced sample tests, protocol was an unexpected moderator of manipulation check scores, *b* = 2.46, 95% CI [2.26, 2.67]. Although the depletion task was more difficult and effortful than the non-depletion task in both protocols, the difference was substantially larger in the writing task protocol compared to the E-task protocol.

We analyzed other task self-reports in a similar manner. The fatigue index revealed higher scores in the depletion condition than in the non-depletion condition. Similarly, reports of frustration were higher among depletion compared to non-depletion participants. Scores on the motivation index again did not differ by condition (Tables S4 and S5).

Exploratory tests of whether the manipulation check reports were moderated by protocol revealed some unanticipated patterns (Table S5). Reports on the effort index were moderated by protocol for both samples. For the reduced sample, that test was preregistered as it comprised the primary check of the manipulation (Table 3 in the main article). Protocol moderated scores on the fatigue index, such that in the writing task protocol, participants in the depletion condition reported being more fatigued than participants in the non-depletion condition, whereas in the E-task protocol, participants in the non-depletion condition reported being more fatigued than participants in the depletion condition. The latter pattern runs contrary to expectations and the published literature (e.g., Baumeister, Bratslavsky, Muraven, & Tice, 1998; Legault, Green-Demers, & Eadie, 2009). Scores on the motivation index also were moderated by protocol. In the writing task protocol, participants in the depletion condition reported being more motivated than did participants in the non-depletion condition, which is another unexpected pattern. Motivation reports did not differ by condition in the E-task protocol. Frustration reports were not moderated by protocol.

We hesitate to speculate about the unexpected patterns for the fatigue and motivation indices, but there may be a few implications. An examination of the conditional means on the fatigue index suggests that the non-depletion task in the E-task protocol was not the clean, neutral exercise we assumed it would be. The motivation index difference, with participants in the controlled writing (versus free writing) condition reporting more motivation, is not consistent with any existing models of the ego depletion effect. The unexpected results from exploratory analyses of the manipulation checks would need to be replicated in future research to bolster confidence in them.

Table S4. *Manipulation Checks: Descriptive Statistics and Exploratory Frequentist Meta-Analytic Tests of Experimental Condition, Full Sample*

Variable	***M* (*SD*)**	**FE Average**	**CI**	**RE Average**	**CI**	**I^2**
Effort index	3.56 (1.74)	1.21**	[1.14,1.29]	1.59**	[1.13, 2.03]	96.88%
Frustration	2.98 (1.94)	0.88**	[0.80, 0.95]	1.01**	[0.70, 1.33]	94.60%
Fatigue index	3.12 (1.56)	0.26*	[0.20, 0.33]	0.27*	[0.12, 0.42]	80.17%
Motivation index	5.23 (1.25)	0.04	[-0.03, 0.11]	0.04	[-0.04, 0.11]	20.84%

Note: *N* = 3528, with the exception that frustration ratings were missing for two participants. Sample size departs slightly from total sample size due to missing data. Condition coded such that 0 = non-depletion, 1 = depletion condition. Higher numbers indicate that participants in the depletion condition reported stronger feelings than participants in the non-depletion condition. All tests were exploratory. *M*s and *SD*s are from unstandardized scales ranging from 1 (*not at all*) to 7 (*very*). FE indicates fixed-effects models; RE indicates random-effects models. CI indicates 95% confidence intervals. * *p*<.05; ** *p*<.01

Table S5. *Manipulation Checks: Descriptive Statistics and Exploratory Frequentist Meta-Analytic Tests of Experimental Condition by Protocol, Full Sample*

			***M* (*SD*)**				**RE Average**		
Variable	**Sample**	**Task**	**Depletion**	**Non-Depletion**	***k***	***N***	***d***	**CI**	**I^2 %**
Effort Index	Reduced	Writing Task	5.81 (1.09)	2.48 (1.03)	16	1246	3.09***	[2.87, 3.30]	39.29
		E-Task	3.27 (1.29)	2.71 (1.18)	20	1217	0.46***	[0.34, 0.57]	0
	Full	Writing Task	5.77 (1.13)	2.52 (1.05)	16	1679	2.98***	[2.71, 3.25]	72.48
		E-Task	3.33 (1.34)	2.74 (1.23)	20	1849	0.45***	[0.36, 0.55]	0
Frustration	Reduced	Writing Task	5.05 (1.65)	1.77 (1.24)	16	1246	2.26***	[2.07, 2.46]	45.04
		E-Task	2.62 (1.55)	2.34 (1.48)	20	1215	0.19**	[0.06, 0.32]	22.08
	Full	Writing Task	4.98 (1.74)	1.89 (1.34)	16	1679	2.01***	[1.84, 2.19]	51.81
		E-Task	2.74 (1.62)	2.42 (1.52)	20	1847	0.19***	[0.10, 0.29]	0
Fatigue Index	Reduced	Writing Task	3.24 (1.59)	2.29 (1.31)	16	1246	0.67***	[0.52, 0.83]	43.08
		E-Task	3.33 (1.47)	3.53 (1.50)	20	1217	-0.15*	[-0.29, -0.01]	30.61
	Full	Writing Task	3.30 (1.61)	2.29 (1.33)	16	1679	0.70***	[0.59, 0.80]	14.61
		E-Task	3.35 (1.51)	3.47 (1.49)	20	1849	-0.10	[-0.20, 0.00]	18.00
Motivation Index	Reduced	Writing Task	4.87 (1.19)	4.62 (1.22)	16	1246	0.19**	[0.07, 0.31]	12.52
		E-Task	5.61 (1.10)	5.70 (1.06)	20	1217	-0.10	[-0.22, 0.01]	1.85
	Full	Writing Task	4.85 (1.22)	4.65 (1.22)	16	1679	0.14**	[0.04, 0.24]	8.10
		E-Task	5.64 (1.11)	5.69 (1.11)	20	1849	-0.06	[-0.15, 0.04]	8.34

Note: Condition coded such that 0 = non-depletion, 1 = depletion condition. Higher numbers indicate that participants in the depletion condition reported stronger feelings than participants in the non-depletion condition. All tests were exploratory. *M*s and *SD*s are from unstandardized scales ranging from 1 (*not at all*) to 7 (*very*). RE indicates random-effects models. CI indicates 95% confidence intervals.
* *p*<.05; ** *p*<.01; *** *p*<.001

Additional Sample and Methodological Details

Recruitment

The lead author (KV) announced the intention to conduct this replication on behavioral science listservs. She also sent personal emails to prominent scholars who have published on ego depletion, including to scholars who have been publicly critical of depletion. Forty laboratories indicated commitment to participating in the project. Six dropped out before initiating or completing data collection and two additional laboratories joined before the end of the data collection period.

Materials and procedures

Participating laboratories received a script for how to conduct the experiment, complete with the wording they should use and the arrangement of the laboratory. When necessary, members of non-English-speaking laboratories translated the script and experimental materials into the language in which the study would be conducted. Additionally, KV created video samples of how to conduct each protocol and shared them with participating labs. Via Skype, KV or BS communicated with laboratories to answer questions and provide additional information. Last, laboratories in both protocols were instructed to have experimenters leave the room while participants performed the study’s tasks (independent variable task, dependent variable task, manipulation check ratings, individual difference measures, demographics, and post-experimental questionnaires).

E-task protocol. The instructions for both pages of this task were in the laboratory’s native language whereas the E-task text was in English for all participants (even if the laboratory’s native language was not English).

Participating laboratories reported the number of errors participants made on the last full paragraph participants completed of the manipulation task used in the E-task protocol. Crossing out an E that should have been skipped and skipping an E that should have been crossed out both counted as errors.

A figure-tracing task served as the dependent measure in this protocol. Experimenters surreptitiously recorded how long participants persisted at figure tracing and counted the number of figure sheets participants attempted to solve.

Story-writing protocol. Laboratories reported uses of forbidden letters (i.e., *a* and *n*) and simple omissions of forbidden letters (e.g., “the dog b_rked”) for each participant in the depletion condition of the story-writing protocol. Only depletion condition participants could have errors. Across both conditions, story word counts were reported.

The CET served as the dependent measure in this protocol. Experimenters timed the duration participants took to complete the CET.

We did not include one item from the published version of the CET, “How much does a telephone weigh?” The published scoring metric (see Bullard et al., 2004; Fein et al., 1998) does not correspond to the weight of contemporary telephones. Additionally, some items on the CET ask for imperial measurements (e.g., “How many sticks of spaghetti are there in a one pound package?”). For labs outside North America, those items were revised to indicate the metric system.

Responses to each item were converted to a common metric before final scoring of the CET. The CET was scored using published norms (Bullard et al., 2004; Fein et al., 1998). Answers within 25-75% of the normative range (i.e., good estimates) received 2 points. Answers outside the 25-75% range but within the 5-95% normative range received 1 point. Answers outside the normative range (i.e., extreme estimates) received 0 points. Participants occasionally gave answers with a tilde (e.g., ~1), which we treated as the numerical value (e.g., 1). Responses given as a range (e.g., 6 to 8) were treated as the median of the two values (e.g., 7).

We considered some answers invalid. Some items did not specify a unit of measurement (e.g., distance could be reported in inches, feet, miles, and so on), and participants were instructed to provide the unit of their response. If they did not provide a unit of measurement for a relevant item, the response was considered invalid. If participants did not report a numerical answer (e.g., "infinite") or provided a nonsensical answer (e.g., "0.5 pounds" for an item asking for a number of spaghetti sticks), the response was considered invalid. Last, if participants skipped an item, it counted as invalid. The final CET score for each participant was an average calculated by summing item scores and dividing by the number of valid responses.

Videos of experimenters. All but two labs submitted recordings of experimenters conducting the study on a practice subject, although five lacked usable audio or video. A total of 65 videos were coded by two independent coders using scales from 1 (*not at all*) to 5 (*very much*) on professionalism (i.e., how competent, in charge, like a leader, and professional in appearance the experimenter behaved), *r* = 0.70, 95% CI [0.68, 0.73], κ = 0.63, *M* = 4.64, *SD* = 0.49), and ease/comfort (i.e., how warm, natural, comfortable, and not stiff or robotic the experimenter behaved), *r* = 0.53, 95% CI [0.50, 0.55], κ = 0.36, *M* = 4.56, *SD* = 0.56). For labs that conducted the study in English, videos (*n* = 49) also were coded for adherence to the script (*r* = 0.72, 95% CI

[0.69, 0.74], κ = 0.44, *M* = 4.61, *SD* = 0.65). The judges' ratings were averaged together and these average scores for professionalism, ease/comfort, and adherence to the script were combined into a composite score of experimenter behavior. Descriptive statistics for the video codings were based on the full sample of participants.

Exclusions

Following preregistered criteria, we excluded data from *n* = 1068 participants as follows. (Some participants failed multiple exclusion criteria.) The overall number of participants who were excluded was more than we expected, but by percentage of all participants the exclusion rate aligns closely with another multi-site depletion replication study. Hagger et al.'s (2016) multi-lab depletion replication paper reported an exclusion rate of 30.9% (*n* = 958 out of 3099 total participants). By comparison, our exclusion rate was 30.25% (1068 out of a total sample size of 3531).

The exclusion criteria can be broadly understood as belonging to four categories: 1) participants' performance errors or mistakes on the tasks (e.g., errors on the E-task, invalid responses on the CET), 2) participants' behavior (e.g., being disturbed, disruptive, or disrupted; using their phone in violation of instructions; knowing that the puzzles were unsolvable in the E-task protocol), 3) participant characteristics (being a non-native speaker of the language in which the study was run; being one of the experimenters' first three participants), and 4) other exclusions. Experimenters noted irregularities that occurred during the course of the study, and three independent coders determined whether each irregularity qualified as an exclusion. (For more information on that process, see below under "Both protocols.") Examples of issues determined to be disqualifying included noise from construction during the study, a repeat participant, missing the timing cue to stop a task, and experimenters being acquainted with participants. Counts of excluded participants based on each preregistered criterion are reported in Table 2 in the main article.

E-task protocol. We excluded data from participants who made more than 2.5 MAD (median absolute deviation) errors on the last full paragraph they completed on the E-crossing task (Leys, Ley, Klein, Bernard, & Licata, 2013). For page 1 of the task (the habit-forming portion), MAD calculations were done at the lab level. For page 2 of the task (the habit-breaking portion), MAD calculations were done within lab and separately by condition. We also excluded data from participants who expressed knowledge (prior to the debriefing) that the figures used in the figure-tracing task (the dependent measure in this protocol) were unsolvable. Table 2 in the main text displays exclusion counts.

Story writing protocol. We excluded data from participants who used 2.5 MAD or fewer words than other participants in their lab and in the same experimental condition, participants who used the restricted letters (*a* and *n*) more often than 2.5 MAD of the lab (this criterion applied only to the depletion condition), and participants who scored beyond 2.5 MAD of the lab mean on invalid responses on the CET (Table 2).

Both protocols. As preregistered, we excluded participants who were non-native speakers as indicated by matching the language(s) they reported speaking at home against the language in which the study was run, who were among the first three run by each experimenter, who reported using their phone during the study, and who were reported by the experimenter to be belligerent, or distressed or distraught. Also as preregistered, we excluded data from participants who experienced a disruption during the experiment session or otherwise experienced an unanticipated deviation from the experimental procedures, as indicated by the experimenter (Table 2).

Further, we instructed experimenters to note other concerns that may warrant excluding the participant. That information was culled and sent to KV, BS, and Rebecca Schlegel, who independently coded whether the concerns merited exclusion of that participant's data. Coders were blind to all other data pertaining to the participant (e.g., condition, protocol, scores on the dependent measures). Exclusions occurred only when all three coders agreed that a participant should be excluded ("Other exclusions;" Table 2). In cases when two of the three coders thought a participant should be excluded, all coders conferred and came to a consensus.

References

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

Bullard, S. E., Fein, D., Gleeson, M. K., Tischer, N., Mapou, R. L., & Kaplan, E. (2004). The Biber cognitive estimation test. *Archives of Clinical Neuropsychology*, *19*, 835-846.

Fein, D., Gleeson, M. K., Bullard, S., Mapou, R., & Kaplan, E. (1998, February). *The Biber Cognitive Estimation Test*. Poster presented at the annual meeting of the International Neuropsychological Society, Honolulu, HI.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*, 546-573.

Legault, L., Green-Demers, I., & Eadie, A. L. (2009). When internalization leads to automatization: The role of self-determination in automatic stereotype suppression and implicit prejudice regulation. *Motivation and Emotion*, *33*, 10-24.

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

APPENDIX A

Full List of Authors

Vohs, Kathleen D., University of Minnesota
Schmeichel, Brandon J., Texas A&M University
Lohmann, Sophie, Max Planck Institute for Demographic Research and University of Illinois at Urbana-Champaign
Gronau, Quentin, F., University of Amsterdam
Finley, Anna, University of Wisconsin-Madison
Ainsworth, Sarah E., Tallahassee Community College
Alquist, Jessica, L., Texas Tech University
Baker, Michael, D., East Carolina University
Brizi, Ambra, University “Sapienza” of Rome
Bunyi, Angelica, University of North Florida
Butschek, Grant, J., University of Georgia
Campbell, Collier, Texas Tech University
Capaldi, Jonathan, Carleton University
Cau, Chuting, University of Toronto
Chambers, Heather, Texas A&M University
Chatzisarantis, Nikos, L. D., Curtin University
Christensen, Weston, J., Brigham Young University-Idaho
Clay, Samuel L., Brigham Young University-Idaho
Curtis, Jessica, Arkansas State University
De Cristofaro, Valeria, University “Sapienza” of Rome
del Rosario, Kareena, University of California, San Francisco
Diel, Katharina, Ruhr University Bochum
Doğruol, Yasemin, Northwestern University
Doi, Megan, University of Minnesota
Donaldson, Tina L., University at Albany
Eder, Andreas B., University of Würzburg
Ersoff, Mia, Florida State University
Eyink, Julie, R., Indiana University
Falkenstein, Angelica, University of California, Riverside
Fennis, Bob M., University of Groningen, the Netherlands
Findley, Matthew, B., Austin College
Finkel, Eli, J., Northwestern University
Forgea, Victoria, Georgia Southern University
Friese, Malte, Saarland University
Fuglestad, Paul, University of North Florida
Garcia-Willingham, Natasha, E., University of Kentucky
Geraedts, Lea F., University of Würzburg
Gervais, Will, M., University of Kentucky
Giacomantonio, Mauro, University “Sapienza” of Rome
Gibson, Bryan, Central Michigan University
Gieseler, Karolin, Saarland University

Gineikiene, Justina, ISM, University of Management and Economics, Vilnius, Lithuania
Gloger, Elana, M., University of Kentucky
Gobes, Carina, M., Florida State University
Grande, Maria, University of Cologne
Hagger, Martin S., University of California, Merced
Hartsell, Bethany, University of North Florida
Hermann, Anthony, D., Bradley University
Hidding, Jasper, J., University of Groningen, the Netherlands
Hirt, Edward R., Indiana University
Hodge, Josh, University of Melbourne
Hofmann, Wilhelm, Ruhr University Bochum
Howell, Jennifer L., University of California, Merced
Hutton, Robert, D., Bradley University
Inzlicht, Michael, University of Toronto
James, Lily, University of Melbourne
Johnson, Emily, Arkansas State University
Johnson, Hannah, L., Brigham Young University-Idaho
Joyce, Sarah, M., Florida State University
Joye, Yannick, ISM University of Management and Economics, Vilnius, Lithuania
Kaben, Jan Helge, Saarland University
Kammrath, Lara, K., Wake Forest University
Kelly, Caitlin, N., Florida State University
Kissell, Brian L., Central Michigan University
Koole, Sander, L., VU Amsterdam
Krishna, Anand, University of Würzburg
Lam, Christine, University of California, Riverside
Lee, Kelemen, T., Bradley University
Lee, Nick, Curtin University
Leighton, Dana, Texas A&M University, Texarkana
Loschelder, David D., Leuphana University Lüneburg
Maranges, Heather, M., Florida State University
Masicampo, E.J., Wake Forest University
Mazara, Jr., Kennedy, Austin College
McCarthy, Samantha, University at Albany
McGregor, Ian, University of Waterloo
Mead, Nicole L., Schulich School of Business, York University
Mendes, Wendy B., University of California, San Francisco
Meslot, Carine, Curtin University
Michalak, Nicholas, M., University of Michigan
Milyavskaya, Marina, Carleton University
Miyake, Akira, University of Colorado Boulder
Moeini-Jazani, Mehrad, University of Groningen, the Netherlands
Muraven, Mark, University at Albany
Nakahara, Erin, University of California, San Francisco
Patel, Krishna, University of Toronto
Petrocelli, John, V., Wake Forest University

Pollak, Katja, M., Leuphana University Lüneburg
Price, Mindi, M., Texas Tech University
Ramsey, Haley, J., Western Kentucky University
Rath, Maximilian, Leuphana University Lüneburg
Robertson. Jacob A., University of Colorado Boulder
Rockwell, Rachael, Ohio University
Russ, Isabella F., University of Würzburg
Salvati, Marco, University “Sapienza” of Rome
Saunders, Blair, University of Dundee
Scherer, Anne, Wake Forest University
Schütz, Astrid, University of Bamberg
Schmitt, Kristin N., University of Colorado Boulder
Segerstrom, Suzanne C., University of Kentucky
Serenka, Benjamin, University at Albany
Sharpinskyi, Konstantyn, University of Waterloo
Shaw, Meaghan, Carleton University
Sherman, Janelle, Indiana University
Song, Yu, Wake Forest University
Sosa, Nicholas, Ohio University
Spillane, Kaitlyn, University of California, Riverside
Stapels, Julia, University of Cologne
Stinnett, Alec, J., Texas Tech University
Strawser, Hannah, R., Texas A&M University
Sweeny, Kate, University of California, Riverside
Theodore, Dominic, Ohio University
Tonnu, Karine, Texas Tech University
van Oldenbeuving, Yasmijn, VU Amsterdam
vanDellen, Michelle R., University of Georgia
Vergara, Raiza, C., Florida State University
Walker, Jasmine, S., East Carolina University
Waugh, Christian, E., Wake Forest University
Weise, Feline, VU Amsterdam
Werner, Kaitlyn, M., Carleton University
Wheeler, Craig, University of Waterloo
White, Rachel, A., East Carolina University
Wichman, Aaron L., Western Kentucky University
Wiggins, Bradford, J., Brigham Young University-Idaho
Wills, Julian A., New York University
Wilson, Janie H., Georgia Southern University
Wagenmakers, E.J., University of Amsterdam
Albarracín, Dolores, University of Illinois at Urbana-Champaign

APPENDIX B
PIs and Laboratory Members

*Ainsworth, Sarah E., Tallahassee Community College
Bunyi, Angelica, University of North Florida
*Fuglestad, Paul, University of North Florida
Hartsell, Bethany, University of North Florida

*Alquist, Jessica, L., Texas Tech University
Campbell, Collier, Texas Tech University
Price, Mindi, M., Texas Tech University
Stinnett, Alec, J., Texas Tech University
Tonnu, Karine, Texas Tech University

*Baker, Michael, D., East Carolina University
Walker, Jasmine, S., East Carolina University
White, Rachel, A., East Carolina University

*Clay, Samuel L., Brigham Young University-Idaho
Christensen, Weston, J., Brigham Young University-Idaho
Johnson, Hannah, L., Brigham Young University-Idaho
*Wiggins, Brady, J., Brigham Young University-Idaho

*Curtis, Jessica, Arkansas State University
Johnson, Emily, Arkansas State University

*Hagger, Martin S., University of California, Merced
Chatzisarantis, Nikos, L. D., Curtin University
Lee, Nick, Curtin University
Meslot, Carine, Curtin University

*Hermann, Anthony, D., Bradley University
Hutton, Robert, D., Bradley University
Lee, Kelemen, T., Bradley University

*Hirt, Edward R., Indiana University
Eyink, Julie, R., Indiana University
Sherman, Janelle, Indiana University

*Howell, Jennifer L., University of California, Merced
Rockwell, Rachael, Ohio University
Sosa, Nicholas, Ohio University
Theodore, Dominic, Ohio University

*Fennis, Bob M., University of Groningen, the Netherlands
Gineikiene, Justina, ISM, University of Management and Economics, Vilnius, Lithuania

Hidding, Jasper, J., University of Groningen, the Netherlands
Joye, Yannick, ISM University of Management and Economics, Vilnius, Lithuania
Moeini-Jazani, Mehrad, University of Groningen, the Netherlands

*Findley, Matthew, B., Austin College
Mazara, Jr., Kennedy, Austin College

*Finkel, Eli, J., Northwestern University
Doğruol, Yasemin, Northwestern University

*Friese, Malte, Saarland University
Kaben, Jan Helge, Saarland University
Gieseler, Karolin, Saarland University

*Giacomantonio, Mauro, University “Sapienza” of Rome
Brizi, Ambra, University “Sapienza” of Rome
De Cristofaro, Valeria, University “Sapienza” of Rome
Salvati, Marco, University “Sapienza” of Rome

*Hofmann, Wilhelm, Ruhr University Bochum
Diel, Katharina, Ruhr University Bochum
Grande, Maria, University of Cologne
Stapels, Julia, University of Cologne

*Inzlicht, Michael, University of Toronto
Cau, Chuting, University of Toronto
Patel, Krishna, University of Toronto
Saunders, Blair, University of Dundee

*Kammrath, Lara, K., Wake Forest University
*Masicampo, E.J., Wake Forest University
*Petrocelli, John, V., Wake Forest University
*Scherer, Anne, Wake Forest University
*Song, Yu, Wake Forest University
*Waugh, Christian, E., Wake Forest University

*Kissell, Brian L., Central Michigan University
Gibson, Bryan, Central Michigan University

*Koole, Sander, L., VU Amsterdam
van Oldenbeuving, Yasmijn, VU Amsterdam
Weise, Feline, VU Amsterdam

*Krishna, Anand, University of Würzburg
Eder, Andreas B., University of Würzburg
Geraedts, Lea F., University of Würzburg

Russ, Isabella F., University of Würzburg

*Leighton, Dana, Texas A&M University, Texarkana

*Loschelder, David D., Leuphana University Lüneburg
Pollak, Katja, M., Leuphana University Lüneburg
Rath, Maximilian, Leuphana University Lüneburg

*Maranges, Heather, M., Florida State University
Ersoff, Mia, Florida State University
Gobes, Carina, M., Florida State University
Joyce, Sarah, M., Florida State University
Kelly, Caitlin, N., Florida State University
Vergara, Raiza, C., Florida State University

*McGregor, Ian, University of Waterloo
Sharpinskyi, Konstantyn, University of Waterloo
Wheeler, Craig, University of Waterloo

*Mead, Nicole L., Schulich School of Business, York University
Hodge, Josh, University of Melbourne
James, Lily, University of Melbourne

*Mendes, Wendy B., University of California, San Francisco
del Rosario, Kareena, University of California, San Francisco
Nakahara, Erin, University of California, San Francisco

*Milyavskaya, Marina, Carleton University
Capaldi, Jonathan, Carleton University
Werner, Kaitlyn, M., Carleton University
Shaw, Meaghan, Carleton University

*Miyake, Akira, University of Colorado Boulder
Robertson, Jacob A., University of Colorado Boulder
Schmitt, Kristin N., University of Colorado Boulder

*Muraven, Mark, University at Albany
Donaldson, Tina L., University at Albany
McCarthy, Samantha, University at Albany
Serenka, Benjamin, University at Albany

*Schmeichel, Brandon J., Texas A&M University
Chambers, Heather, Texas A&M University
Finley, Anna, University of Wisconsin-Madison
Strawser, Hannah, R., Texas A&M University

*Schütz, Astrid, University of Bamberg

*Segerstrom, Suzanne C., University of Kentucky
Gloger, Elana, M., University of Kentucky
Garcia-Willingham, Natasha, E., University of Kentucky

*Sweeny, Kate, University of California, Riverside
Lam, Christine, University of California, Riverside
Spillane, Kaitlyn, University of California, Riverside
Falkenstein, Angelica, University of California, Riverside

*vanDellen, Michelle R., University of Georgia
Butschek, Grant, J., University of Georgia

*Wichman, Aaron L., Western Kentucky University
Ramsey, Haley, J., Western Kentucky University

*Wilson, Janie H., Georgia Southern University
Forgea, Victoria, Georgia Southern University

Note: Laboratories are listed under the name of the PI used in the tables and figures, followed by additional members. For ease of presentation, tables and figures refer to each laboratory using the name of a PI, although some groups had more than one PI. The Wake Forest laboratory considered all members to be PIs and therefore is listed by site.

* indicates laboratory PIs.