Research Article

Psychological Science
1–16

DOI: 10.1177/0956797621989733
www.psychologicalscience.org/PS
SAGE

A Multisite Preregistered Paradigmatic Test of the Ego-Depletion Effect

Kathleen D. Vohs[1], Brandon J. Schmeichel[2], Sophie Lohmann[3,4], Quentin F. Gronau[5], Anna J. Finley[6], Sarah E. Ainsworth[7], Jessica L. Alquist[8], Michael D. Baker[9], Ambra Brizi[10], Angelica Bunyi[11], Grant J. Butschek[12], Collier Campbell[8], Jonathan Capaldi[13], Chuting Cau[14], Heather Chambers[2], Nikos L. D. Chatzisarantis[15], Weston J. Christensen[16], Samuel L. Clay[16], Jessica Curtis[17], Valeria De Cristofaro[10], Kareena del Rosario[18], Katharina Diel[19], Yasemin Doğruol[20], Megan Doi[1], Tina L. Donaldson[21], Andreas B. Eder[22], Mia Ersoff[23], Julie R. Eyink[24], Angelica Falkenstein[25], Bob M. Fennis[26], Matthew B. Findley[27], Eli J. Finkel[20], Victoria Forgea[28], Malte Friese[29], Paul Fuglestad[11], Natasha E. Garcia-Willingham[30], Lea F. Geraedts[22], Will M. Gervais[31], Mauro Giacomantonio[10], Bryan Gibson[32], Karolin Gieseler[29], Justina Gineikiene[33], Elana M. Gloger[30], Carina M. Gobes[23], Maria Grande[34], Martin S. Hagger[35,36], Bethany Hartsell[11], Anthony D. Hermann[37], Jasper J. Hidding[26], Edward R. Hirt[24], Josh Hodge[38], Wilhelm Hofmann[19], Jennifer L. Howell[35], Robert D. Hutton[37], Michael Inzlicht[14], Lily James[39], Emily Johnson[17], Hannah L. Johnson[16], Sarah M. Joyce[23], Yannick Joye[40], Jan Helge Kaben[29], Lara K. Kammrath[41], Caitlin N. Kelly[23], Brian L. Kissell[32], Sander L. Koole[42], Anand Krishna[22], Christine Lam[25], Kelemen T. Lee[37], Nick Lee[15], Dana C. Leighton[43], David D. Loschelder[44], Heather M. Maranges[23], E. J. Masicampo[41], Kennedy Mazara, Jr.[27], Samantha McCarthy[21], Ian McGregor[45], Nicole L. Mead[46], Wendy B. Mendes[47], Carine Meslot[15], Nicholas M. Michalak[48], Marina Milyavskaya[13], Akira Miyake[49], Mehrad Moeini-Jazani[26], Mark Muraven[21], Erin Nakahara[47], Krishna Patel[14], John V. Petrocelli[41], Katja M. Pollak[44], Mindi M. Price[8], Haley J. Ramsey[50], Maximilian Rath[44], Jacob A. Robertson[49], Rachael Rockwell[51], Isabella F. Russ[22], Marco Salvati[10], Blair Saunders[52], Anne Scherer[41], Astrid Schütz[53], Kristin N. Schmitt[49], Suzanne C. Segerstrom[30], Benjamin Serenka[21], Konstantyn Sharpinskyi[45], Meaghan Shaw[13], Janelle Sherman[24], Yu Song[41], Nicholas Sosa[51], Kaitlyn Spillane[25], Julia Stapels[34], Alec J. Stinnett[8], Hannah R. Strawser[2], Kate Sweeny[25],

Dominic Theodore[51], Karine Tonnu[8], Yasmijn van Oldenbeuving[42], Michelle R. vanDellen[12], Raiza C. Vergara[23], Jasmine S. Walker[9], Christian E. Waugh[41], Feline Weise[42], Kaitlyn M. Werner[13], Craig Wheeler[45], Rachel A. White[9], Aaron L. Wichman[50], Bradford J. Wiggins[16], Julian A. Wills[18], Janie H. Wilson[28], Eric-Jan Wagenmakers[54], and Dolores Albarracín[4]

[1]Department of Marketing, Carlson School of Management, University of Minnesota; [2]Department of Psychological and Brain Sciences, Texas A&M University; [3]Max Planck Institute for Demographic Research, Rostock, Germany; [4]Department of Psychology, University of Illinois at Urbana-Champaign; [5]Department of Social and Behavioural Sciences, University of Amsterdam; [6]Center for Healthy Minds, University of Wisconsin-Madison; [7]Behavioral Sciences Department, Tallahassee Community College; [8]Department of Psychological Sciences, Texas Tech University; [9]Department of Psychology, East Carolina University; [10]Department of Psychology, Sapienza University of Rome; [11]Department of Psychology, University of North Florida; [12]Department of Psychology, University of Georgia; [13]Department of Psychology, Carleton University; [14]Department of Psychology, University of Toronto; [15]School of Psychology, Curtin University; [16]Department of Psychology, Brigham Young University-Idaho; [17]Department of Psychology & Counseling, Arkansas State University; [18]Department of Psychology, New York University; [19]Faculty of Psychology, Ruhr University Bochum; [20]Department of Psychology, Northwestern University; [21]Department of Psychology, University at Albany; [22]Department of Psychology, University of Würzburg; [23]Department of Psychology, Florida State University; [24]Department of Psychological and Brain Sciences, Indiana University; [25]Department of Psychology, University of California, Riverside; [26]Department of Marketing, University of Groningen; [27]Psychology Department, Austin College; [28]Department of Psychology, Georgia Southern University; [29]Department of Psychology, Saarland University; [30]Department of Psychology, University of Kentucky; [31]Centre for Culture and Evolution, Psychology, Brunel University London; [32]Psychology Department, Central Michigan University; [33]Department of Marketing, ISM University of Management and Economics; [34]Department of Psychology, University of Cologne; [35]Psychological Sciences, University of California, Merced; [36]Faculty of Sport and Health Sciences, University of Jyväskylä; [37]Psychology Department, Bradley University; [38]School of Psychological Sciences, University of Melbourne; [39]London College of Fashion, University of the Arts London; [40]Department of Management, ISM University of Management and Economics; [41]Department of Psychology, Wake Forest University; [42]Department of Clinical Psychology, Vrije Universiteit Amsterdam; [43]College of Arts, Sciences, and Education, Texas A&M University, Texarkana; [44]Institute of Psychology, Leuphana University Lüneburg; [45]Department of Psychology, University of Waterloo; [46]Schulich School of Business, York University; [47]Department of Psychiatry and Behavioral Sciences, University of California, San Francisco; [48]Department of Psychology, University of Michigan; [49]Department of Psychology and Neuroscience, University of Colorado Boulder; [50]Department of Psychology, Western Kentucky University; [51]Psychology Department, Ohio University; [52]School of Social Sciences, University of Dundee; [53]Department of Psychology, University of Bamberg; and [54]Department of Psychology, University of Amsterdam

Abstract

We conducted a preregistered multilaboratory project (k = 36; N = 3,531) to assess the size and robustness of ego-depletion effects using a novel replication method, termed the *paradigmatic replication approach*. Each laboratory implemented one of two procedures that was intended to manipulate self-control and tested performance on a subsequent measure of self-control. Confirmatory tests found a nonsignificant result (d = 0.06). Confirmatory Bayesian meta-analyses using an informed-prior hypothesis (δ = 0.30, SD = 0.15) found that the data were 4 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 exclusion criteria) found a statistically significant effect (d = 0.08); Bayesian analyses showed that the data were about equally likely under the null and informed-prior hypotheses. Exploratory moderator tests suggested that the depletion effect was larger for participants who reported more fatigue but was not moderated by trait self-control, willpower beliefs, or action orientation.

Keywords

ego depletion, self-control, registered replication, open data, open materials, preregistered

Received 12/3/19; Revision accepted 11/24/20

Corresponding Author:

Kathleen D. Vohs, University of Minnesota, Carlson School of Management, Department of Marketing

E-mail: kvohs@umn.edu

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, self-control being 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, so using self-control on an initial task renders subsequent self-control less successful than if not deployed earlier (Baumeister et al., 1998; Muraven et al., 1998).

The concept of ego depletion has been widely influential. The seminal article on the subject (Baumeister et al., 1998) has had "transformational" impact (Nosek et al., 2010, supplement). In addition to inspiring a multitude of empirical articles, the theory inspired multiple new theories as well (e.g., Evans et al., 2016; Inzlicht & Schmeichel, 2012; Job et al., 2010; for a review, see Baumeister & Vohs, 2016b). 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 multisite preregistered test of the phenomenon. Challenges to ego depletion have come in two main forms: meta-analytic analyses (Carter et al., 2015) and a multisite registered replication study (Hagger et al., 2016). Those investigations cast doubt on depletion theory but have been criticized on methodological and analytical grounds (Baumeister & Vohs, 2016a; Friese et al., 2019; Garrison et al., 2019; Inzlicht et al., 2015). Germane to the current study is that the previous replication study used methods uncommon to depletion studies (for a rebuttal, see Hagger & Chatzisarantis, 2016). As a result, we conducted a multisite 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 was intended to create a new model for replication studies. Termed the *paradigmatic replication approach*, it made multiple changes to existing models (for details on how the current project differs from others, see Spellman & Kahneman, 2018). 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. Researchers at the 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 (K. D. Vohs) 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 data set and blinded the data set[1] before sending it on to the analysis team. The analysis team conducted preregistered analyses before sharing results with the lead authors (K. D. Vohs and B. J. Schmeichel), who then generated recommendations for exploratory analyses. Lead authors had access to the data only after the analysts had done their work (see Table 1).

Experimental Protocols

Each laboratory 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 et al., 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, as compared with people who did not use self-control initially. We expected the magnitude of the effect to be equivalent across protocols (see preregistration at https://osf.io/952mv/).

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

Goal	Strategy	Rationale
	Formulation stage	
Identify representative tasks	Crowdsource with area experts: create list of possible independent-variable and dependent-variable tasks deemed paradigmatic for testing the hypothesis.	Collect diversity of possible methods and 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 and 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, and consider using both frequentist and Bayesian approaches.	Use open-science practices, expand skill set beyond what project leaders bring, and increase information value of results.
	Study-preparation stage	
Public statements of intent	Preregister hypotheses, methods, and participant exclusion criteria and specify conclusions given different possible results.	Use open-science practices and reduce researcher degrees of freedom.
Methods testing and practice	Make video recordings of how to conduct the study; write and review scripts for experimenters to follow.	Reduce variation in procedural execution.
Team building	Conduct virtual meetings with all members of participating labs.	Address questions, reinforce procedural details, and bridge gaps between project leaders and data-collection labs.
	Post-data-collection stage	
Ensure data integrity	Labs send data to independent handler who (a) merges data files, (b) blinds outcome measures, and (c) sends master data set 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 and perform tests that were unanticipated or underspecified in preregistration.

We chose manipulation checks common to the depletion literature, namely, participants' reports of the difficulty of the initial task, the degree of effort required for it, and feelings of frustration from it (Hagger et al., 2010). Other self-report measures included reports of being tired or fatigued. We predicted that compared with people in the nondepletion 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 condition than in the nondepletion condition. The original 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 by 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.[2] 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).

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 et al.,

Table 2. Exclusion Counts for Each Preregistered Criterion

Criterion	E-task protocol	Writing-task protocol
Errors on last completed E-task paragraph (p. 1)	159	
Errors on last completed E-task paragraph (p. 2)	133	
Knew puzzles were unsolvable	42	
Used few words in story		7
Used forbidden letters in story		83
Invalid responses on Cognitive Estimation Test		0
Nonnative 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
Total	823	527

Note: In total, 1,068 participants were excluded in accordance with preregistered criteria. Some participants (n = 237) failed multiple exclusion criteria. For additional details, see the Supplemental Material available online.

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 (e.g., DeWall et al., 2007) trait self-control; 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, number of depletion studies published by principal investigators, and laboratory location (see the Supplemental Material available online).

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

Method

Participants

Thirty-six laboratories (see the Supplemental Material) tested 3,531 people (2,375 women, 1,130 men, 11 who listed "other," and 15 who did not report gender; age: M = 20.92 years, SD = 5.19). Most of the laboratories were located in the United States (k = 23), and there were 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 = 1,679), and 20 laboratories chose the E-task protocol (n = 1,852). Among all participants, 1,762 were randomly assigned to the depletion condition, and 1,769 were randomly assigned to the nondepletion condition. On the basis of preregistered criteria, we excluded 30.25% (n = 1,068) 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 with other multisite replications, see Table 2 and the Supplemental Material). 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 Supplemental Material.

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. For each of the operationalizations, those scholars provided feedback on 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 the writing-task protocol; we assigned those two 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/). The Supplemental Material contains additional details.

Experimental procedures

Overview. Both protocols followed the same basic procedure, 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 postexperimental questionnaire (https://osf.io/952mv/).

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 (nondepletion 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 min for the first page and 8 min 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. After ensuring that participants understood, experimenters laid down stacks of the three test images, telling participants that they could quit the task at any time 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 that they were done with the task (or after 20 min).

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 = .39, 95% confidence interval [CI] = [.35, .43]).

Writing-task protocol. Participants' first task was to write a story about a recent trip. Participants in the nondepletion condition received no additional instructions. Participants in the depletion condition were further instructed not to use words containing the letters "A" and "N" in their story. Participants wrote for 5 min in both conditions. After the writing task, the experimenter introduced the dependent measure, the 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 (see the Supplemental Material), we averaged the points to form a final CET score, which was then standardized.

Manipulation checks. After completing 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 ranging 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 = .71). A sample item is, "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-oriented option they choose). 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 = .81). A sample item is, "I am good at resisting temptation" (1 = *not at all like me*, 5 = *very much like me*). Last, participants completed the six-item Strenuous Mental Activity subscale of the Implicit Theories About Willpower Scale[3] (Job et al., 2010), which measures whether people think that self-control is a limited resource (M = 4.18, SD = 0.90; α = .84; n = 2,452). A sample item is, "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 so 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).[4] Advisory board members provided invaluable help in formulating hypotheses, suggesting analytical models, analyzing data, and preregistering the project.

Data-set procedures. After labs completed data collection, they sent a data set 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 (K. D. Vohs and B. J. Schmeichel) would not have access to the data until after the analysts[5] 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 sending the data set to the analysts. Thus, lead authors did not have access to the data until after the analysts did, nor did they conduct analyses. After the initial analyses were conducted, the data set was unblinded. In 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 meta-analyses on each laboratory's Cohen's d effect size, representing the difference between the nondepletion and depletion conditions. (Fixed-effects analyses are reported in parentheses.) Larger effect sizes indicate a stronger depletion effect (i.e., lower scores on the dependent measures of self-control). Analyses were conducted in the R programming environment (Version 4.0.2; R Core Team, 2020). Moderators were tested using multilevel linear models in the individual-level analyses (with the *lme4* package; Version 1.1-26; Bates et al., 2015) and using random-effects meta-regression for meta-analytic analyses at the lab level (with the *metafor package*; Version 2.4-0; Viechtbauer, 2010).

Bayesian statistics. Bayes factors (BFs) addressed the evidentiary basis of the depletion effect. To address the question of whether the effect exists, we pitted a point-null hypothesis, which states that the effect is absent, against an informed one-sided alternative hypothesis centered on a depletion effect (δ) of 0.30 with a standard deviation of 0.15. The estimate for the preregistered alternative hypothesis 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 (d) of 0.62, and Hagger et al.'s (2016) Registered Replication Report, which reported an overall effect size of 0.04. We split the difference and arrived at a δ of 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 BFs (e.g., Jeffreys, 1939) to quantify the relative support for the informed hypothesis over the point-null hypothesis.

Subsequent analyses provided information on the size of the ego-depletion effect after the data from the study were taken into account. 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-effects and random-effects models according to their plausibility given the data (Gronau et al., 2017; Scheibehenne et al., 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-studies standard deviation (van Erp et al., 2017).

Results

In this section, we report preregistered and thus confirmatory analyses on the reduced sample (i.e., after excluding participants on the basis of preregistered criteria; see Table 2). First, we report results on the

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

			Fixed-effects models		Random-effects models		
Variable	Depletion condition (*M*)	Nondepletion condition (*M*)	Average Cohen's *d*	95% CI	Average Cohen's *d*	95% 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: $N = 2,463$ ($k = 36$), with the exception that frustration ratings were missing for two participants. Standard deviations are given in parentheses. Sample size varies from total sample size because of missing data. Condition were coded as 0 (nondepletion condition) and 1 (depletion condition). Higher numbers indicate that participants in the depletion condition reported stronger feelings than participants in the nondepletion condition. All tests were confirmatory (preregistered). Means and standard deviations are from unstandardized scales (1 = *not at all*, 7 = *very*). CI = confidence interval.
$^{**}p < .01$. $^{***}p < .001$.

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

Exploratory tests can be found in the Supplemental Material. 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 (i.e., with no participants excluded).

Manipulation checks

Frequentist analyses. Meta-analyses were conducted to check the effectiveness of the depletion task (see 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 nondepletion condition. Although scores on the effort index showed substantial heterogeneity across laboratories (*d*s ranged from 0.08, 95% CI = [−0.65, 0.81], to 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% CI = [1.66, 1.86], $I^2 = 0\%$) represented the average effect across both protocols (−0.5 = E-task, 0.5 = writing task). We did not expect protocol to moderate scores on the effort index but preregistered our intention to 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 nondepletion 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 nondepletion 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 coefficient = .90) and, as per our preregistration, were averaged to form an index of fatigue. As predicted, the main effect of depletion condition was significant: Participants reported more fatigue in the depletion than the nondepletion condition. Also as expected, participants in the depletion condition reported feeling more frustrated than did participants in the nondepletion condition (see Table 3; see also Tables S4 and S5 in the Supplemental Material).

Reports of motivation and wanting to do well on the dependent measure formed an internally consistent scale (Spearman-Brown coefficient = .74). The two items were standardized and averaged to form a motivation index. We preregistered competing predictions: (a) that there would be no depletion-condition effect

Table 4. Depletion Effect: Frequentist Meta-Analyses

		Random-effects meta-analysis			Fixed-effects meta-analysis	
Dependent variable	*n*	*d*	95% CI	I^2 (%)	*d*	95% CI
Overall depletion effect	2,461	0.06	[−0.02, 0.14]	2.54	0.06	[−0.02, 0.14]
Overall figure-tracing performance	1,216	0.12	[−0.01, 0.24]	15.16	0.11	[−0.00, 0.22]
Figure-tracing duration	1,216	0.15*	[0.02, 0.29]	28.46	0.13*	[0.02, 0.25]
Figure-tracing attempts	1,217	0.05	[−0.06, 0.17]	0	0.05	[−0.06, 0.17]
Cognitive Estimation Test	1,245	0.01	[−0.10, 0.12]	0	0.01	[−0.10, 0.12]

Note: Sample sizes vary because of missing data. For analyses of the overall depletion effect, k = 36; for figure-tracing analyses, k = 20; and for Cognitive Estimation Test analyses, k = 16. Conditions were coded as 0 (nondepletion condition) and 1 (depletion condition). Higher numbers indicate evidence of a depletion effect (i.e., that self-control was worse in the depletion condition). CI = confidence interval.
*p < .05.

(in line with the depletion theory) or (b) that participants in the depletion condition would report lower motivation than participants in the nondepletion condition (in line with the model proposed by Inzlicht & Schmeichel, 2012). Consistent with the former prediction, results showed no difference in self-reported motivation (see Table 3; see Tables S4 and S5).

Bayesian analyses. To quantify the predictions under Hypothesis 1 (H1), we conducted a model-averaged Bayesian meta-analysis using a one-sided Cauchy prior on effect size μ with mode 0 and scale 0.707. Given that the preregistration plans for the primary outcome variable specified using a β(1, 2) prior distribution for the between-study heterogeneity τ, we adopted that approach here. However, in work succeeding the preregistration, we have consistently used an inverse-γ prior with a shape of 1 and a scale of 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 β prior and for the inverse-γ prior. Noticeable differences between these priors are due to the fact that the β prior does not allow values for τ higher than 1, contrary to what the data suggested.

For the effort index, BF(β prior) was greater than 1.797693×10^{308} and BF(inverse-γ prior) equaled 1,123,563; for feelings of frustration, BF(β prior) equaled 2,727,844,064 and BF(inverse-γ prior) equaled 85,152; and for the fatigue index, BF(β prior) equaled 5.68 and BF(inverse-γ prior) equaled 6.13. For motivation, both priors yielded the same BF in favor of the null hypothesis (BF_{+0}) of 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 they provide moderate evidence that depletion increased feelings of fatigue. The depletion manipulations did not affect self-reported motivation.

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 nondepletion conditions was not statistically significant, d = 0.06, 95% CI = [−0.02, 0.14] (see Table 4 and Fig. 1).

Bayesian analyses. The presence of a depletion effect was then tested using Bayesian analyses. In these analyses, a BF_{+0} of 10 would indicate that the data are 10 times more likely under the informed alternative hypothesis, which is centered on a δ of 0.30, than under the point-null hypothesis. Correspondingly, a BF_{+0} of 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 BFs quantified the overall evidence in favor of either the informed alternative hypothesis or the point-null hypothesis across all laboratories simultaneously. The meta-analytic BF of focal interest was the model-averaged one (see Fig. 2). For comparison, we displayed the meta-analytic BF for the fixed-effects and random-effects models separately. All three meta-analytic BFs showed close agreement and favored the point-null hypothesis to approximately the same degree (see Fig. 2). The model-averaged BF indicated that the data were 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 (see Fig. 3).[6] This BF value indicates moderate evidence in favor of the point-null hypothesis according to the classification scheme proposed by Jeffreys (1939).

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

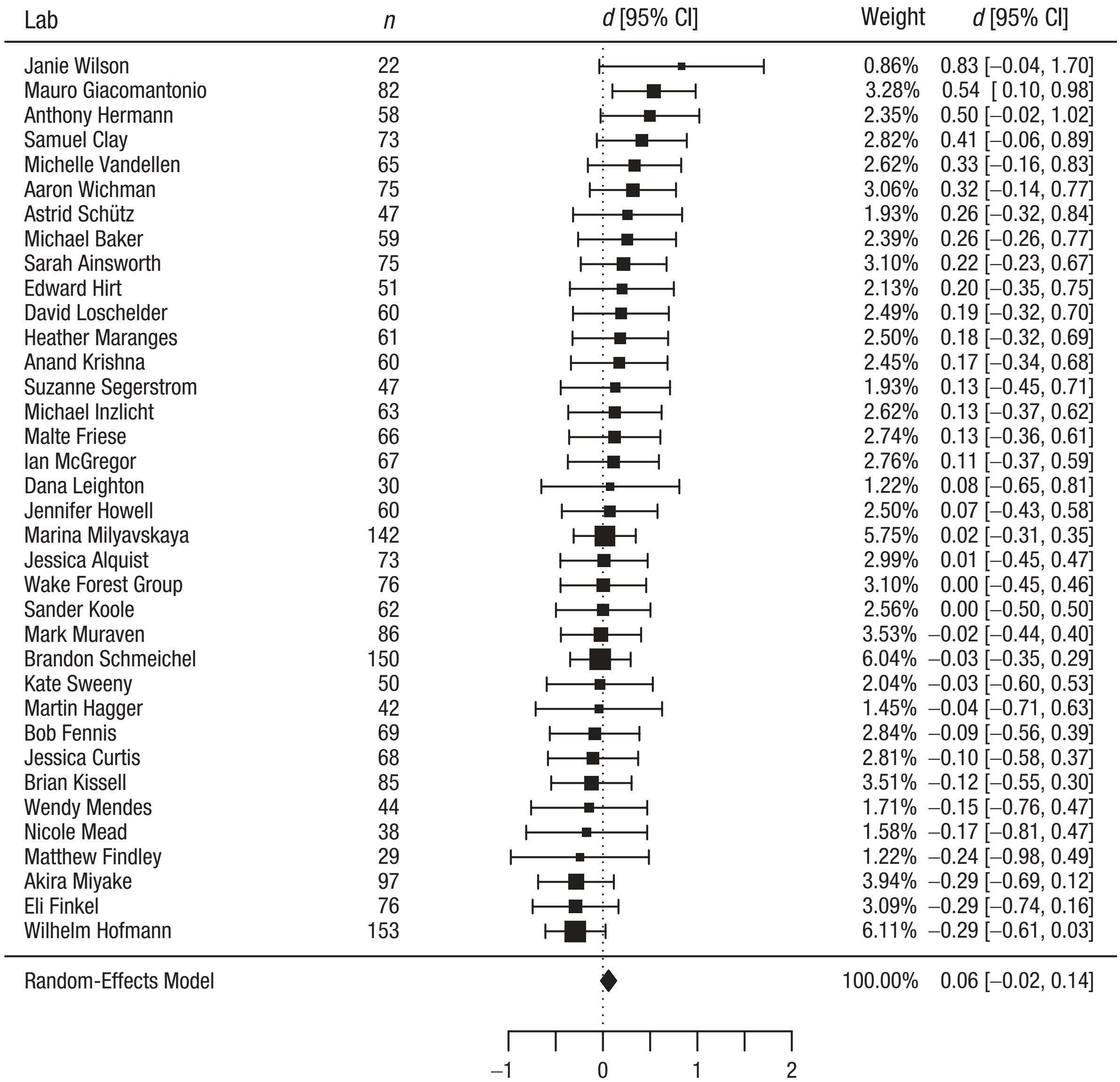

Fig. 1. Forest plot of performance outcome by laboratory. The box plots and numerical values illustrate the same effect-size estimates. For the plots, the size of each box represents its weighted contribution to the overall effect, and its whiskers display the 95% confidence interval (CI). The dotted line represents a zero effect size. Numerical effect sizes show standardized mean differences between the depletion and nondepletion conditions. The diamond is the overall meta-analytic effect derived from a random-effects model. Laboratories are referred to by the name of a principal investigator (PI), although some labs had more than one PI. The Wake Forest laboratory considered all members to be PIs and therefore is listed by site.

negative effect-size values. When examining individual laboratories' data, we found that 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* of 0.30.

We inspected the data (assuming a nonzero effect) across individual laboratories, but they 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 a δ of 0.30 toward zero. The posterior median was 0.08, 95% CI = [0.01, 0.16] (see Fig. 3).

Potential moderators

We first checked whether outcomes varied by protocol (the specific combination of manipulation and dependent measures). The dependent measure was performance,

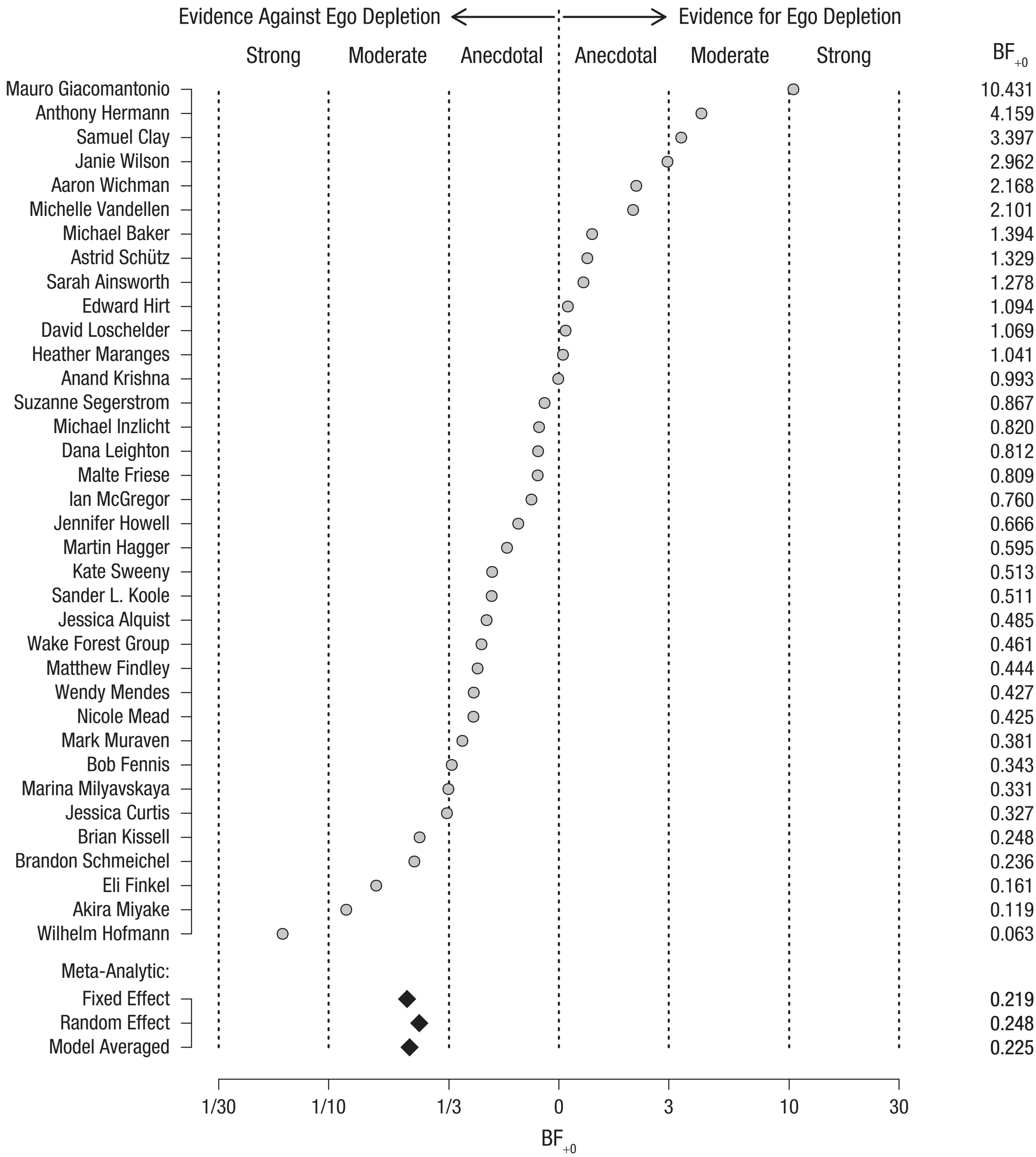

Fig. 2. Bayesian forest plot of performance outcome by laboratory. The values listed under BF_{+0} indicate relative support for the depletion hypothesis versus a hypothesis that there was no effect. Diamonds indicate overall effect sizes from meta-analytic models using fixed effects and random effects and one that combined both approaches. Laboratories are referred to by the name of a principal investigator (PI), although some labs had more than one PI. The Wake Forest laboratory considered all members to be PIs and therefore is listed by site. BF = Bayes factor.

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 in random-effects 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 significant moderator, suggesting that the magnitude of the effect did not differ across protocols.

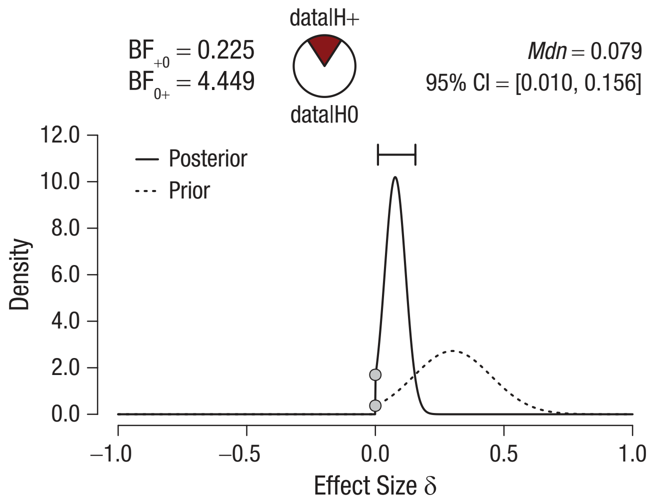

Fig. 3. Tests of the model-averaged meta-analytic effect-size posterior and Bayes factor. The dotted line indicates the informed-prior effect-size distribution, and the solid line indicates the model-averaged meta-analytic posterior effect-size distribution. Roughly speaking, the peak of the shape indicates the likelihood of the effect size, and its width indicates variance. The two gray dots on the plot are the prior and posterior ordinates at delta = 0 (the prior being the lower dot). The pie chart shows a visual representation of the relative support for the alternative hypothesis (H+) and the null hypothesis (H0). CI = confidence interval.

The total score on the figure-tracing task was the combination of the number of puzzle sheets that 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 nonsignificant effect of condition (see Table 4).

We preregistered our intention to separately examine 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 et al., 2009; Vohs et al., 2008). In the current data, however, the two figure-tracing components exhibited only a moderate correlation (r = .39, 95% CI = [.35, .43]).

Examining the two components separately, we found that the effect of condition on number of attempts was not statistically significant (unstandardized descriptive statistics: nondepletion condition: M = 19.87, SD = 9.92; depletion condition: M = 19.36, SD = 10.41; see Table 4).

In contrast, there was a significant effect on duration (random-effects model: d = 0.15, 95% CI = [0.02, 0.29]; fixed-effects model: d = 0.13, 95% CI = [0.02, 0.25]; see Table 4). Participants in the depletion condition gave up about 27 s sooner on the figure-tracing task than participants in the nondepletion condition (unstandardized descriptive statistics: nondepletion condition: M = 1,012.20 s, SD = 266.30; depletion condition: M = 985.10 s, 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 Supplemental Material. 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 with the nondepletion) condition among participants who reported being more fatigued by the manipulation task (see Table S2 and Fig. S4 in the Supplemental Material).

Discussion

We tested an ego-depletion hypothesis on more than 3,500 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.[7] Confirmatory frequentist analyses indicated that the two conditions did not differ, although outcome performance was directionally worse in the depletion condition compared with the nondepletion condition (d = 0.06; see Table 4). Confirmatory Bayesian tests found more evidence for the absence than presence of a depletion effect (see Fig. 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; see Table S1 and Fig. S1 in the Supplemental Material) depletion effect. Comparable Bayesian analyses showed no clear evidence to support or refute the informed alternative hypothesis in support of a depletion effect (see Figs. S2 and S3 in the Supplemental Material).

Moving back to frequentist tests, the findings suggested that self-reported fatigue acted as a moderator. The more that participants in the depletion condition felt fatigued, the worse their subsequent self-control (see Table S2 and Fig. S4). This pattern is congruent with prior evidence regarding the role of subjective fatigue in depletion effects (e.g., 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 (see Table 4 and Figs. 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 (δ) falls between 0.15 and 0.45, fit the full-sample data about equally well (see Figs. 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 (see Table S1 and Fig. S1). Exploratory analyses showing significant moderation by self-reported fatigue further suggest that depletion effects may be conditional (see Table S2 and Fig. 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, the E-task protocol showed a bigger depletion effect (in terms of descriptive statistics) than the writing-task protocol.[8] 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 who reported 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 et al., 2014) may be worthwhile.

The current project was inspired by a previous multilab study of ego depletion, which reported an effect size of $d = 0.04$ (Hagger et al., 2016). A recent multilab test reported an effect size of $d = 0.10$. That study tested the same individual differences as did we, finding little in the way of moderation (Dang et al., 2021).

All told, the results from two multilab investigations compare similarly with 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 multilab 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 (i.e., were paradigmatic) while allowing for 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—to help determine 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 multisite 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 were 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 (see Table 1).

The goals of these implementations were twofold. 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 an intention to 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 depletion research (and beyond) are 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: A larger effect was observed among participants in the depletion condition who reported greater fatigue. Readers doubtful of the theory may see the findings as damning for the 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 sizes) 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.

Transparency

Action Editor: D. Stephen Lindsay
Editor: D. Stephen Lindsay
Author Contributions
K. D. Vohs and B. J. Schmeichel developed the study concept and design. K. D. Vohs and B. J. Schmeichel wrote the manuscript with help from S. Lohmann, D. Albarracín, Q. F. Gronau, and E.-J. Wagenmakers, who also analyzed and interpreted the data. All of the authors approved the final manuscript for submission.
Declaration of Conflicting Interests
The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.
Funding
This research was supported by Netherlands Organisation for Scientific Research grants to Q. F. Gronau (406.16.528) and to E.-J. Wagenmakers (016.Vici.170.083) as well as an Advanced European Research Council grant to E.-J. Wagenmakers (743086 UNIFY).
Open Practices
All data and materials have been made publicly available via OSF and can be accessed at https://osf.io/952mv. The design and analysis plan were preregistered on AsPredicted, and copies can be viewed at https://osf.io/952mv. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iDs

Kathleen D. Vohs https://orcid.org/0000-0001-5858-2603
Sophie Lohmann https://orcid.org/0000-0001-9852-7684
Quentin F. Gronau https://orcid.org/0000-0001-5510-6943
Eric-Jan Wagenmakers https://orcid.org/0000-0003-1596-1034

Acknowledgments

We thank Roy Baumeister, Eddie Harmon-Jones, Rebecca Schlegel, Uri Simonsohn, Jake Westfall, and Wendy Wood for assistance.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797621989733

Notes

1. This method of blinding the data involves the data handler switching the labels of key hypothesis-testing variables with other variables so as to reduce the potential for bias in the analysis process (Dutilh et al., 2019).
2. 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 *nondepletion condition* refers to a task designed to require relatively less self-control.
3. Because of formatting errors, some laboratories omitted the Implicit Theories About Willpower Scale, resulting in different sample sizes.
4. The advisory board members were D. Albarracín, W. M. Gervais, Q. F. Gronau, S. Lohmann, E.-J. Wagenmakers, J. Westfall, and W. Wood.
5. The analysts were D. Albarracín, Q. F. Gronau, S. Lohmann, and E.-J. Wagenmakers.
6. Our recent work uses an inverse-γ distribution, which we applied to the confirmatory depletion-hypothesis test. Results did not appreciably change compared with those using the β distribution (see Fig. 3). Using an inverse-γ prior for between-study heterogeneity τ, we found that the model-averaged meta-analytic BF_{+0} equaled 0.228 or, expressed in favor of the null, BF_{0+} equaled 4.39.
7. This language reflects our preregistered conclusion if analyses showed a nonsignificant result.
8. By referring to protocols by their manipulation tasks, we do not mean to imply that those tasks necessarily made the difference. The dependent measures may have been differentially sensitive or other factors may have been at work.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

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). Misguided effort with elusive implications. *Perspectives on Psychological Science*, *11*(4), 574–575. https://doi.org/10.1177/1745691616652878

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

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 1155. https://doi.org/10.3389/fpsyg.2016.01155

Dang, J., Barker, P., Baumert, A., Bentvelzen, M., Berkman, E., Buchholz, N., Buczny, J., Chen, Z., De Cristofaro, V., de Vries, L., Dewitte, S., Giacomantonio, M., Gong, R., Homan, M., Imhoff, R., Ismail, I., Jia, L., Kubiak, T., Lange, F., . . . Zinkernagel, A. (2021). A multilab replication of the ego depletion effect. *Social Psychological and Personality Science*, *12*(1), 14–24. https://doi.org/10.1177/1948550619887702

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.

Dutilh, G., Sarafoglou, A., & Wagenmakers, E.-J. (2019). Flexible yet fair: Blinding analyses in experimental psychology. *Synthese*. Advance online publication. https://doi.org/10.1007/s11229-019-02456-7

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. R., Boggero, I. A., & Segerstrom, S. C. (2016). The nature of self-regulatory fatigue and "ego depletion": Lessons from physical fatigue. *Personality and Social Psychology Review*, *20*(4), 291–310. https://doi.org/10.1177/1088868315597841

Fein, D., Gleeson, M. K., Bullard, S., Mapou, R., & Kaplan, E. (1998, February). *The Biber Cognitive Estimation Test* [Poster presentation]. International Neuropsychological Society, Honolulu, HI, United States.

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*(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*, Article 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. R., Brand, R., Brandt, M. J., Brewer, G., Bruyneel, S., Calvillo, D. P., Campbell, W. K., Cannon, P. R., Carlucci, M., Carruth, N. P., Cheung, T., Crowell, A., De Ridder, D. T. D., Dewitte, S., . . . Zwienenberg, M. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science*, *11*(4), 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., & Berkman, E. (2015). *Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Commentary on Carter, Kofler, Forster, & McCullough, 2015*. SSRN. http://dx.doi.org/10.2139/ssrn.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*(5), 450–463.

Jeffreys, H. (1939). *Theory of probability*. 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*(11), 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). 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., Frazier, R., & Tenney, E. R. (2010). Cumulative and career-stage citation impact of social-personality psychology programs and their members. *Personality and Social Psychology Bulletin*, *36*(10), 1283–1300.

R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.2) [Computer software]. R Foundation for Statistical Computing.

Scheibehenne, B., Gronau, Q. F., Jamil, T., & Wagenmakers, E.-J. (2017). Fixed or random? A resolution through model-averaging: Reply to Carlsson, Schimmack, Williams, and Burkner. *Psychological Science*, *28*(11), 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. *Behavioral and Brain Sciences*, *41*, Article 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*(6), 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., & 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*(1), Article 4. http://doi.org/10.5334/jopd.33

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3). https://doi.org/10.18637/jss.v036.i03

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.