A Multi-Lab Pre-Registered Replication of the Ego-Depletion Effect

Proposing authors: Martin S. Hagger and Nikos L. D. Chatzisarantis
Contributing authors: Hugo Alberts, Calvin Octavianus Anggono, Angela Birt, Ralf Brand, Mark J. Brandt, Gene Brewer, Sabrina Bruyneel, Dustin Calvillo, W. Keith Campbell, Tracy Cannon, Nicholas Carruth, Tracy Cheung, Adrienne Crowell, Denise T. D. De Ridder, Siegfried Dewitte, Malte Elson, Jaqueline Evans, Benjamin A. Fay, Bob M. Fennis, Anna Finley, Zoë Francis, Elke Heise, Henrik Hoemann, Michael Inzlicht, Sander L. Koole, Lina Koppel, Floor Kroese, Florian Lange, Kevin Lau, Bridget P. Lynch, Carolien Martijn, Akira Miyake, Alexandra Mosser, Megan Muise, Dominique Muller, Milena Muzzi, Dario Nalis, Ratri Nurwanti, Henry Otgaar, Michael Phillipp, Pierpaolo Primoceri, Katrin Rentzsch, Lara Ringos, Caroline Schlinkert, Brandon J. Schmeichel, Sarah F. Schoch, Michel Schrama, Astrid Schütz, Angelos Stamos, Gustuv Tinghög, Johannes Ullrich, Michelle vanDellen, Supra Wimbarti, Wanja Wolff, Cleoputri Yusainy, Oulmann Zerhouni, Maria Zweinenberg

Protocol vetted by: Roy Baumeister, Chandra Sripada, Daniel Kessler
Edited by: Alex O. Holcombe

Multi-lab direct replication of: Ego-depletion paradigm reported in Sripada, C., Kessler, D., & Jonides, J. (2014). Methylphenidate blocks effort-induced depletion of regulatory control in healthy volunteers. *Psychological Science, 25*(6), 1227-1234. doi:10.1177/0956797614526415.

Data and registered protocols: https://osf.io/jymhe/
Commentary:

Citation: Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Birt, A., Brand, R., … Zweinenberg, M. (2015). A Multi-Lab Pre-Registered Replication of the Ego-Depletion Effect. *Perspectives on Psychological Science,* X, X-X.

Abstract

The conceptualization of self-control as a limited resource is a popular model. The resource-depletion effect has typically been tested using a dual-task experimental paradigm. As a result of completing a task requiring self-control, people are predicted to have reduced self-control capacity and therefore poorer performance on a subsequent task,, a state known as 'ego-depletion'. While one recent meta-analysis of ego-depletion experiments found a medium-sized effect, others have questioned the size and existence of the effect and identified instances of possible publication bias. The analyses served as a catalyst for the current registered replication report (RRR) of the ego-depletion effect. Multiple laboratories (N = 24) conducted replications of a standardized ego-depletion protocol adopting the dual-task paradigm and computer-administered tasks: an initial letter 'e' task, with depletion and no depletion versions, and a subsequent multi-source interference task (MSIT). Reaction time variability (RTV) and overall reaction time (RT) on the MSIT constituted dependent measures of ego-depletion. Meta-analysis of the labs' data revealed small RTV (d = .020) and RT (d = -.031) effects with 95% confidence intervals that encompassed zero. Results support recent meta-analytic findings suggesting null findings for the ego-depletion effect. We discuss implications of the findings for the ego-depletion effect and the resource depletion model of self-control.

Key words: strength model; energy model; resource depletion; self-regulation; meta-analysis

A Multi-Lab Pre-Registered Replication of the Ego-Depletion Effect

In two key research articles, Baumeister and colleagues (Baumeister, Bratslavsky, Muraven, & Tice, 1998; Muraven, Tice, & Baumeister, 1998) proposed and tested a 'limited resource' or 'strength' model of self-control. According to their model, performance on tasks requiring self-control is governed by a generalized, unitary, and finite 'internal' resource. They proposed that engaging in tasks requiring self-control would lead to the depletion of the resource and reduced performance on subsequent self-control tasks. The state of reduced self-control capacity was termed 'ego-depletion'. Baumeister and colleagues tested their model using a 'dual task' experimental paradigm in which participants engaged in two consecutive tasks. For participants randomly allocated to the experimental (ego-depletion) group, both tasks required self-control while for participants allocated to the control (no depletion) group only the second task required self-control, with the first task not requiring any, or very little, self-control. The self-control tasks used typically required participants to alter or modify an instinctive, well-learned response, akin to resisting an impulse or temptation (Baumeister, Vohs, & Tice, 2007). As predicted, participants in the experimental group performed worse on the second task relative to participants in the control group. Critically, the tasks used in the experiments were from different 'domains' of self-control providing evidence to suggest that the resource was 'domain-general' and common to all tasks that required self-control. The limited resource account has received considerable support with numerous replications of the original findings using the dual-task paradigm. An initial meta-analysis revealed a medium effect size (d = 0.62) across 198 tests of the ego-depletion effect (Hagger, Wood, Stiff, & Chatzisarantis, 2010).

There have, however, been recent conceptual and empirical analyses that have challenged the resource depletion explanation for the ego-depletion effect and questioned the strength of the effect or whether it exists at all. Recent analyses have suggested that original

meta-analytic effect size for ego-depletion may be inflated or even zero. Re-analyses of previous meta-analytic findings (Carter & McCullough, 2013b; Carter & McCullough, 2014) and a new meta-analysis of tests of the ego-depletion effect including some unpublished data (Carter, Kofler, Forster, & McCullough, 2015) applied regression techniques based on funnel plots of the estimated effect size in each study against study precision (i.e., the reciprocal of the sample size). Analyses revealed substantial 'small study' bias in the effect size reported in Hagger et al.'s (2010) original meta-analysis and that many published studies included in the original analysis, and in the revised analysis, were substantially underpowered suggesting that the likelihood of finding the large, significant effects reported in the literature was improbable. In both their re-analysis and updated meta-analysis, Carter et al. (2015) suggested that, based on their regression analyses, that a probable value for the ego-depletion effect was zero and concluded that "the meta-analytic evidence does not support the proposition (and popular belief) that self-control functions as if it relies on a limited resource, at least when measured as it typically is in the laboratory" (p. 18). Consistent with these findings, there have also been studies that have failed to replicate the ego-depletion effect (e.g., Xu et al., 2014), found it to be substantially smaller in size than reported in meta-analytic syntheses (e.g., Tuk, Zhang, & Sweldens, 2015), or have even indicated that a facilitation effect may occur in multi-task experiments involving self-control tasks (e.g., Converse & DeShon, 2009; Tuk et al., 2015). Overall, these data, together with the data from the recent meta-analyses, cast doubt on the existence of a large or even moderately-sized ego-depletion effect.

It is important, however, to note that the interpretation of the regression analyses conducted by Carter et al. has been questioned. Hagger and Chatzisarantis (2014) indicated that the initial analyses were problematic given the limitations of interpreting bias identified using the regression techniques when substantial heterogeneity in the effect sizes exists across studies. Furthermore, the regression techniques are based on an assumption that the

relationship between sample size and effect size is zero, but Simonsohn and colleagues (2009) point to instances where this may not be the case (e.g., cases where there is considerable unexplained heterogeneity in the effect size or the sample may have been selected based on a characteristic making them more prone to the depletion manipulation). Importantly, while the regression techniques may identify instances of bias, they cannot provide an explanation for the source of the bias (Simonsohn, 2009).

Issues of interpretation notwithstanding, the issue of substantial bias across studies testing the ego-depletion effect is important and the size of the effect is still uncertain given competing interpretations of tests of bias of the meta-analytic findings. The literature on the ego-depletion effect is a reflection of broader current debates over the reproducibility of effects in psychological experiments (Pashler & Harris, 2012) and the need for high-powered, replications of prominent effects in the discipline (Open Science Collaboration, 2012, 2015). We proposed a set of independent replications of the ego-depletion effect using the dual-task paradigm, as advocated by Carter and McCullough (2013b; 2014) and Hagger and Chatzisarantis (2014).

Protocol Development

The development of the ego-depletion replication protocol was conducted in consultation with Roy Baumeister, the lead author of the original experiments on ego-depletion and a major proponent of the resource-depletion model of self-control. While the dual-task paradigm has become the primary means by which to test the ego-depletion effect, there is considerable variation in the tasks used in the literature, due to researchers' desire to demonstrate the domain generality of the self-control 'resource'. For example, 'exerting' self-control on a task in one domain (e.g., impulse control) was expected to lead to observed decrements in performance on tasks from another (e.g., thought or emotion suppression). A

consequence of this variability in task selection means that there is no single agreed standardized set of tasks for use in dual-task tests of the ego-depletion effect.

A further issue in developing the protocol was the need for tasks to be sufficiently standardized to avoid, wherever possible, idiosyncratic lab-specific differences in the presentation of tasks or other variation that may reduce the consistency of the protocol implementation across labs. While typical practice in registered replications of psychological research has tended to prioritize the replication of the original experiment, the tasks used in the original experiments were deemed too elaborate or involved to be appropriate for a multi-lab replication. For example, the tasting of cookies over radishes task (Baumeister et al., 1998, Study 1) would require extensive experimenter involvement in its delivery. Similarly, persistence on unsolvable anagrams (Baumeister et al., 1998, Study 3) may be too culture-specific, and it would be difficult to develop equivalence in the anagrams across labs from different countries. Furthermore, we also considered it appropriate to adopt 'behavioral' tasks after Carter and colleagues' (2015) plea for researchers to do so in their meta-analysis. We therefore sought to identify a dual-task procedure that adopted standardized 'behavioral' tasks requiring little adaptation across labs and minimal interpersonal involvement by the experimenter.

We adopted the tasks and procedure developed and published by Sripada, Kessler, and Jonides (2014). The tasks used are similar to those used in the original depletion experiments (e.g., Baumeister et al., 1998; Muraven et al., 1998), but computer-administered, a design feature that minimizes variability across labs. The decision was made in close consultation with Roy Baumeister, and with Chandra Sripada and Daniel Kessler, co-authors of the original experiment, who made the same tasks and procedure used in the original study available for the replication project. Note that Sripada et al.'s original study adopted a 2 x 2 design which included methylphenidate supplementation and placebo conditions alongside the ego-depletion

protocol, so the procedure adopted in the current replication is not a direct replication of Sripada et al.'s study but instead a test of the ego-depletion effect adopting their depletion paradigm.

Once the protocol was finalized, a public announcement of the replication and a call for participating labs was posted by *Perspectives on Psychological Science* on October 28, 2015. A deadline for applications to participate was set for January 9, 2015 and by that time 30 labs' applications had been approved by the editor to conduct a replication. Six laboratories had to abort their data collection due to technical difficulties or insufficient resources (e.g., access to participants or research assistance) leaving 24 labs contributing to the project. Participating labs conducted independent replications and pre-registered their implementation plan on the Open Science Framework. Each implementation plan was vetted by the editor for consistency with the protocol prior to data collection. Participating labs were in Australia, Belgium, Canada, France, Germany, Indonesia, the Netherlands, Sweden, Switzerland, and the United States. Details of how materials were translated from English are provided below.

Co-ordinated and systematic translation efforts were undertaken to prepare study materials in labs recruiting participants whose native language was not English. The lead investigators for the participating labs have expertise in social psychology, social cognition, self-regulation and self-control, or experimental design and are listed as co-authors on this manuscript. Some labs had no previous experience in conducting studies on self-control but had expertise in conducting psychology experiments.

Protocol Requirements

In this section we provide details of the replication protocol. From this general protocol, participating labs were required to create an entry on the Open Science Framework (OSF) linked from the main ego-depletion Sripada et al. registered replication report (RRR) webpage (https://osf.io/jymhe/) and post their implementation plan, registration documents, materials,

raw data, and analyses. The study protocol was required to be approved by labs' institutional review board (IRB) or the equivalent institutional committee responsible for research ethics in advance of data collection.

Participants

Participants were recruited from an undergraduate participant pool for course credit or undergraduate students participating for payment. Given evidence suggesting that older participants show a weaker ego depletion effect, participants were required to be between 18 and 30 years old (Dahm et al., 2011). Given that the original task materials were in English and one of the self-control tasks in the proposed dual-task paradigm was language-sensitive, participants in the American, Australian, Canadian, UK, and New Zealand labs were excluded if they did not report English as their first and primary language. The study materials were translated into other languages to permit labs from non-English speaking countries participating in the replication. In each of those labs, participants were excluded if they were not native or fluent speakers of the primary language in which the replication was conducted[1].

Testing Location

Participants were tested individually in laboratory conditions and were alone in the room when completing the tasks. Participants were provided with written instructions and were guided orally by the experimenter, who followed a script.

Experimenters

Researchers were postgraduate psychology students, research assistants, postdoctoral researchers, or faculty researchers with experience in collecting psychology experimental data and interacting with participants. Experimenters did not need to have specific domain knowledge or prior familiarity with the paradigm. Experimenters were required to familiarize

[1]One lab conducted the replication on fluent English-speaking students in Sweden.

themselves with the experimental step-by-step procedure available on the OSF site (https://osf.io/ifdj3/) and practice it prior to data collection. The protocol recommended that experimenters be naïve to the experimental hypothesis and condition assignment, but this was not always feasible (whether it was attempted is noted on each lab's OSF page).

Data Collection

The one-way experimental design reflected Sripada and colleagues' (2014) depletion paradigm after omitting the methylphenidate condition. Participants were allocated to experimental (ego-depletion) or control (no depletion) groups pseudo-randomly. In order to achieve approximately equal numbers of participants across conditions and achieve the minimum numbers required, participants were randomized in blocks of 10 to ensure both conditions met the minimum required sample size. As a result, one condition may have had more participants than the other due to different rates of exclusion, but both would meet the required minimum.

Procedure

The experiment was presented as an experiment on "word and number recognition and reaction time" to mask the study hypothesis. The detailed procedure is posted at https://osf.io/ifdj3/. Participants were welcomed by the experimenter, shown into the lab, and asked to sit at a desk with a computer. They were informed that they would be required to engage in two computer-administered tasks, presented consecutively, after a period of practice on each task. Participants then completed practice versions of the two tasks. The practice versions of both tasks were conducted prior to the main trials in order to minimize transition time between the initial and second tasks in the depletion paradigm. The first task was the letter 'e' task and the second task was the modified multi-source interference task (MSIT, detailed below in the "Materials" section). Both tasks were presented on a computer screen controlled by *E-Prime* experimental software. After the practice sessions, participants proceeded to the

main trials of each task. After completing the first task, task effort, fatigue, difficulty, and frustration was assessed via self-report and used as manipulation checks for the depletion manipulation. Participants then completed the second task. In an exit questionnaire, participants' thoughts about the purpose of the experiment were probed. They were then thanked and informed that the experiment had concluded. Some of the participating labs' IRB required experimenters to provide participants with a debrief. Some labs were able to delay the debrief until completion of the experiment and all participants data had been collected to minimize potential for the study hypothesis being shared with others in the participant pool. Other provided a debrief immediately after the experiment but asked participants not to share details with fellow students.

Materials

Letter 'e' task. The first task was a modified version of the letter 'e' task (Baumeister et al., 1998, Study 4) with on-screen instructions administered using *E-Prime* experimental presentation software (available at https://osf.io/ifdj3/). Two versions were used: 'regulation' and 'control'. In the regulation version administered to the experimental (depletion) group, participants were presented with a series of words on a video screen and required to press a button when a word with the letter 'e' is displayed, but must withhold the response if the 'e' is next to or one letter away from a vowel. The control version administered to the control (no depletion) group was matched in all respects with the exception that participants were required to press a button whenever a word with the letter 'e' is displayed, with no stipulation to ever withhold their response to an 'e'. Participants were asked to respond as quickly and accurately as possible. Participants completed some practice trials before the commencement of the experiment. The main session comprised 150 trials and lasted 7 minutes 30 seconds. Participants' reaction times and errors for the letter 'e' task were recorded by the *E-Prime* program.

Multi-source interference task (MSIT). The MSIT is a task requiring response inhibition (Bush, Shin, Holmes, Rosen, & Vogt, 2003) and was administered by *E-Prime* (available at https://osf.io/ifdj3/). Numeric stimuli were presented on the computer screen with participants making responses using the keyboard. Participants were asked to place their index, middle and ring fingers of the right hand on three keys on the keyboard. Participants were told that they will be presented with sets of three digits (1, 2, 3 or 0) in the center of a video screen every few seconds, and that one digit (the target digit) would always be different from the other two (matching distractor) digits. Participants were told that they needed to indicate the *identity* of the target digit, not its *position* in the set of digits. Participants were required to press the key corresponding to the digit that differed from the other two. In 'control' sets, the target digit (1, 2 or 3) always matched its position on the response keys, such as the number '1' appearing in the first (leftmost) position. For example, sets 100, 020 or 113 are examples of 'control' sets. In 'interference' or 'incongruent' number sets, the target number (1, 2 or 3) never matched its position, and the distractors were themselves potential targets. For example, for the number set 233 the correct response would be '2'. The task creates interference in that the identity of the target number and its position relative to other numbers on the string differed. Interference is also caused by varying the size of digits in the set. In the congruent version the size always matched the target digit while in the incongruent version digit size was not matched to the target, requiring the participant to inhibit both the position and size of the target number in favor of its identity. Participants were asked to respond as quickly and accurately as possible. The task lasted approximately 10 minutes and comprised 200 trials (100 control (congruent) and 100 interference (incongruent) trials) presented in an interspersed, pseudorandom order. Reaction time and error data were recorded by the E-Prime program.

Performance on the MSIT comprised the dependent measure of regulatory performance. The MSIT provides two performance measures: mean reaction time (RT) on incongruent trials

and reaction time variability (RTV) on incongruent trials, defined as the sum of the sigma and tau variability parameters using ex-Gaussian modeling (Dawson, 1988; Sripada et al., 2014). RTV on the MSIT was the primary dependent variable in Sripada et al.'s (2014) original study and in the replication protocol. Attentional control maintains consistent task-directed focus and supresses task-irrelevant spontaneous thoughts. Reduction in attentional control induced by depletion is likely to lead to more lapses in attention, manifesting as increased variability in response latencies across trials on the MSIT (Weissman, Roberts, Visscher, & Woldorff, 2006).While this should also inflate mean RT, the variability is a more sensitive measure. We also conducted analyses on (mean) RT on MSIT incongruent items as a secondary dependent variable as this is the typical criterion variable in other commonly-used interference tasks such as the Stroop color-naming task.

Translation for non-English speaking labs. Labs collecting data from non-English speaking countries were required to translate all study materials into their native language by a fluent bi-lingual translator followed by back-translation by an independent fluent bilingual translator. The specific translation procedure for each lab is documented on their OSF webpage. Assistance in developing the non-English word stimuli and instruction slides for the letter 'e' task and embedding them into the *E-Prime* program was provided by Daniel Kessler, who developed the original tasks in the Sripada et al. (2014) study. The analysis plan was to conduct one meta-analysis of the data from all the participating labs, plus separate meta-analyses for English and non-English-speaking labs.

Data Stopping rules and Exclusions

Each lab pre-registered their stopping rules for data collection, how they planned to meet the demographic requirements of the participants, how they would assess the first and primary language of participants, how participants would be assigned to conditions, and rules for exclusion of participants' data from the analysis. The editor reviewed these procedures to

verify that participating labs lab met protocol requirements. Participant exclusion criteria were decided on before data collection. The criteria were: the participant reported that their native language was one other than the language in which the experiment was conducted, they fell outside the stipulated 18 to 30 years of age, they did not complete the study, they did not follow instructions (e.g., failing to understand instructions), or their responses fell below the 80% correct response criteria for the letter 'e' or MSIT tasks. Participants were also excluded due to equipment or software failure or experimenter error. Raw data files that include data that were excluded from analysis are provided on participating labs' OSF pages (https://osf.io/jymhe/).

Critical comparisons

By convention in dual-task paradigm studies examining the ego-depletion effect, the critical analysis is a one-way test of difference on task performance across the dependent and control groups. Accordingly for the present study, the critical test is comparison of MSIT performance for the experimental (ego-depletion) group to that of the control (no depletion) group. The primary dependent variable was RTV for incongruent trials of the MSIT. The critical test was whether RTV was higher for participants receiving the regulation version of the letter 'e' task (depletion) condition than for those in the no regulation letter 'e' task (control) condition. This is identical to the critical tests conducted in the replicated experiment (Sripada et al., 2014). It is also consistent with the critical tests in the original ego-depletion experiments (Baumeister et al., 1998; Muraven et al., 1998) and those in the extant literature. In terms of predictions, most labs predicted a non-trivial effect size. Some labs (n = 12) predicted that the replicated effect would be similar in size to that reported in previous meta-analyses (Hagger et al., 2010) or the original study (Sripada et al., 2014), and some (n = 9) indicated it would be lower than reported in previous analyses, but greater than $d = 0.15$.

Additional analyses were planned on secondary dependent variables: mean RT for incongruent items on the MSIT, performance on the letter 'e' task, and self-report measures of effort, difficulty, fatigue, and frustration. Larger RTs among participants assigned to the depletion group relative to participants assigned to the control group would be indicative of an ego-depletion effect. Better accuracy and greater levels of effort, difficulty, fatigue and frustration in the regulation version of the letter 'e' task condition relative to the control version would indicate that the regulation version was more arduous and effortful relative to the control version.

Results

Lab Demographics and Results

Sample demographics and results for each participating lab are provided in Table 1 for the depletion and non-depletion conditions. The table provides sample sizes, details of exclusions and reasons, and the means and standard deviations of three of the dependent variables in each condition. Demographics for all participating labs can be found in Appendix A.

Data Analyses: Original and Present

In Sripada et al.'s original study, a two-way analysis was conducted examining the interactive effect of the depletion manipulation and methylphenidate administration conditions on the dependent variables. In the current analysis, consistent with convention in ego-depletion experiments, our critical comparison was test of difference (independent samples *t*-test) for the primary and secondary dependent variables, mean RTV and RT for incongruent items on the MSIT task, respectively, across the depletion conditions. We supplemented this with identical analyses of overall accuracy on the letter 'e' task and participants' self-reports of effort, fatigue, difficulty, and frustration to establish the extent to which the initial task likely involved

effortful self-control. Each lab conducted these analyses independently and results are reported on their OSF project webpages (https://osf.io/jymhe/).

Effect Size Measurements

Differences in the key dependent variables across conditions in pooled data from the labs were tested using meta-analysis for each dependent variable. We used a random effects model to weight each effect by its sample size and report the effect size in standard deviation units (Cohen's *d*) and its confidence intervals. Heterogeneity in the effect sizes was evaluated using the Cochrane *Q* and I^2 statistics, with a significant value for Q and an I^2 value greater than 25% indicative of substantial heterogeneity in the effect size across studies. Forest plots showing the means and standard deviations in both conditions for each lab, the effect size measured in each lab with 95% confidence intervals, and the sample-weighted meta-analytic effect size for the dependent variables of interest (mean RTV and RT for the MSIT) are provided in Figures 1 and 2. Positive effect sizes for RT and RTV represent the extent of a deficit in performance on the second task and an ego-depletion effect while negative numbers go against the effect. Plots and effect size data for each lab for the letter 'e' task accuracy and participants' scores on effort, fatigue, difficulty, and frustration scales and results are presented in Appendix B. Summary statistics from the meta-analyses for all dependent variables are presented in Table 2.

Averaged sample-weighted effect sizes for the mean RTV (d = 0.020, 95% confidence interval: -0.089 to 0.130) and RT (d = -0.031, 95% confidence interval: -0.115 to 0.053) variables were small and confidence intervals included the value of zero. Only three of the 24 replications did not have 95% confidence intervals that included zero for RTV, and one of those was negative. Similarly, only four labs found mean RT values with confidence intervals that did not include the value of zero, three of which were negative effects. We also found substantial heterogeneity in the effect sizes for mean RTV (I^2 = 39.651%, Q_{23} = 38.111, p = .025) and RT (I^2 = 68.374%, Q_{23} = 72.724, p < .001) indicating substantial variability in the

effect across labs. This finding suggests the presence of moderator variables across laboratories, despite them all running the experiment with very strict inclusion criteria and an identical study protocol. Given that every laboratory observed only very small effect sizes for both dependent variables, it is unlikely that a moderator analysis would return a substantive or statistically significant effect size, but it may serve to resolve the heterogeneity.

A candidate moderator identified a priori was the language of the participants. As planned we conducted separate meta-analyses on the data from English speaking and non-English speaking labs. Results are shown in the subgroup analyses in Figures 1 and 2. While there were only very small observed differences in effect sizes in the English speaking labs' data for mean RTV (d = 0.092, 95% confidence interval: -0.068 to 0.252) and RT (d = -0.126, 95% confidence interval: -0.347 to 0.095) variables relative to the non-English speaking labs (RTV: d = -0.042, 95% confidence interval: -0.191 to 0.107; RT: d = -0.007, 95% confidence interval: -0.219 to 0.206), the moderator analysis served to produce homogenous cases in both the English speaking (I^2 = 43.64%, Q_{11} = 19.518, p = .052) and non-English speaking (I^2 = 34.817%, Q_{11} = 16.875, p = .112) labs for mean RTV. The analysis also produced a homogenous case for the non-English speaking labs (I^2 = 20.376%, Q_{11} = 13.185, p = .243), but not the English speaking labs (I^2 = 80.797%, Q_{11} = 57.283, p < .001), for RT. The reductions in the heterogeniety notwithstanding, there was substantial overlap in the confidence intervals of each moderator group and all encompassed zero as a possible value.

Meta-analysis of participants' accuracy on the letter 'e' task and self-report ratings of effort, fatigue, difficulty, and frustration are presented in Appendix B (see Figures B1-B5). We found large effects for depletion condition on letter 'e' task accuracy (d = 0.779, 95% confidence interval: 0.630 to 0.929), and scores on effort (d = 0.779, 95% confidence interval: 0.630 to 0.929), difficulty (d = 1.919, 95% confidence interval: 0.1.714 to 2.123), and frustration (d = 0.815, 95% confidence interval: 0.665 to 0.966), but a substantially smaller

effect for fatigue with confidence intervals that included zero (d = 0.090, 95% confidence interval: -0.018 to 0.197). Overall, these findings provide some evidence that the regulation version of the letter 'e' task administered in the ego-depletion condition was more effortful and aversive than the non-regulation version used in the control group.

Discussion

The current report presents the first registered replication of the ego-depletion effect. Results across all laboratories revealed small effect sizes for the ego-depletion effect on the primary and secondary dependent variables, reaction time variability (RTV) and mean reaction time for incongruent items on the MSIT. In addition, the 95% confidence intervals for the effect sizes for the majority of laboratories' replications included the value of zero. The effects are substantially lower than the ego-depletion effect size for RTV in the placebo condition of Sripada et al.'s (2014) study adopting the current methods (d = 0.69), the uncorrected ego-depletion effect sizes reported in Hagger et al.'s original meta-analysis (d = 0.62) and Carter and colleagues' (2015) revised meta-analysis that comprised 41% unpublished studies (g = 0.43), and bias-corrected meta-analytic estimates such as Carter et al.'s trim-and-fill analysis (g = 0.24). The overall effect size of the replication closely mirrors the regression-based estimate using the precision effect estimation with standard error technique reported by Carter et al. (g = 0.003). Overall, the replicated effect seems to suggest a null effect for ego-depletion for the current paradigm. Although there was substantial heterogeneity in the effect size across laboratories, much of this was resolved when native language of participants was taken into account. Although the moderator analysis revealed modest differences in the effect across English speaking and non-English speaking labs, this variation had little impact on the overall size of the effect and the 95% confidence intervals for the effect for both dependent variables in both groups encompassed zero.

An important issue in depletion experiments using the dual task paradigm is whether the initial task is sufficiently demanding to evoke a depletion effect. From the perspective of the limited resource theory that underpins the ego-depletion effect, the issue is whether the initial task depletes self-control sufficiently to impair performance on the second task. Although there is no direct evidence for this, indication of the extent of depletion after the first task can be determined from measures that tap the extent to which participants invested effort on the first task. In the current replication, performance on the letter 'e' task and self-report measures indicated that the regulation version of the task was more demanding and evoked greater perceptions of effort, difficulty, and frustration than the non-regulation version. This evidence provides some indication that the initial task was more demanding for participants allocated to the depletion condition relative to controls. However, it must be stressed that these data do not provide absolute conformation that the initial task was sufficient to evoke depletion.

Does the current replication suggest that the ego-depletion effect does not exist after all? Certainly the current evidence does raise considerable doubts given the proximity of the tests to the standard dual-task paradigm adopted to test the ego-depletion effect in the literature, and the tightly-controlled tasks and protocol across multiple laboratories. Evidence from the current replication effort suggests that effect sizes observed in many tests of the depletion effect in the literature, including initial meta-analytic estimates may be inflated. In a recent commentary, Inzlicht, Gervais, and Berkman (2015) after Reed et al. (2015) identified a range of estimates of meta-analytic effects derived from different estimation methods including the regression-based estimates reported by Carter et al., an additional estimate that corrects for study bias conducted by Inzlicht et al. based on the *p*-curve (Simonsohn, Nelson, & Simmons, 2014), and the effect sizes derived from the studies with the top ten largest sample sizes in the meta-analyses and the effect size from Carter et al.'s (2015) meta-analysis that includes unpublished studies. Considering the variation in the estimates from the different sources, a definitive

indication of the true ego-depletion effect remains elusive. However, adding the averaged effect size from the current analysis as an additional data point in this portfolio would appear to indicate that, at the very least, the bias-uncorrected effect size estimates derived from meta-analyses is likely to be substantially inflated. Furthermore, given the rigor with which the current replication has been conducted, substantial weight should be attributed to its findings in such considerations.

It would be remiss not to acknowledge a number of limitations that may affect interpretation of the effect size generated in current analysis. First, tasks adopted in the current replication do not conform directly to tasks used in previous dual-task experiments. For example, the regulation version of the letter 'e' task used in the current analysis administered to participants in the ego-depletion condition did not include an initial period where individuals familiarize themselves with the 'easy' or non-regulation version of the task (i.e., the same version used in the control group), prior to engaging the depletion period on the regulation version. The initial period is supposed to induce a 'habitual' response that participants would need to override when engaging in the more demanding regulation version (e.g., Fennis, Janssen, & Vohs, 2009). However, other than Sripada et al.'s original study, a number of dual-task experiments in the literature reported using a letter 'e' task without an initial 'habit forming' period and found depletion effects (e.g., Baumeister et al., 1998; Wan & Sternthal, 2008) and there are also variations of this task (such as Carter and McCullough's (2013a) essay writing task without letters 'a' and 'n') with no initial period. Tasks such as the letter 'e' task with complex rules and time pressure that requires a search for a letter and then making a rule-based decision on whether or not to respond will require the suppression of a tendency for an immediate response. The use of a task without a 'habit forming' period is unlikely to have been a decisive factor in determining whether or not ego-depletion was induced.

Second, it may be that the letter ‘e’ task may have been sufficiently arduous but not of sufficient duration to deplete individuals’ self-control resources. This is consistent with some preliminary evidence that task duration moderates the ego-depletion effect (Hagger et al., 2010). In the current replication, the duration of the letter ‘e’ task was identical to that used by Sripada et al. who found it sufficient to induce depletion. Furthermore, an initial task duration of eight minutes is typical in dual-task experiments. Based on this evidence, we do not think that the current tasks can be dismissed on grounds that they do not sufficiently conform to tasks that have been previously used in depletion experiments.

Finally, an important consideration when evaluating the evidence for the ego-depletion effect is that the effect has been tested in multiple experiments using an array of different initial and dependent tasks in the dual task paradigm. This is consistent with the underlying hypothesis that self-control performance is governed by a generalized resource that is domain general. In other words, engaging in a task in one domain of self-control such as impulse suppression will lead to impaired performance on a task in another domain such as thought or emotional control. Replication of the effect using a particular paradigm in two impulse control tasks, therefore, does not provide definitive evidence for the ego-depletion effect across multiple domains. Coordinated replication efforts adopting multiple tasks from different domains may be needed to provide a definitive account of the depletion effect across multiple domains, a position that has been advocated elsewhere (Carter et al., 2015).

Conclusion

Results from the current multi-lab registered replication of the ego-depletion effect provide evidence that, if there is any effect, is close to zero. When looking at the converging evidence from meta-analyses for the effect, including those that correct for bias, evidence seems to suggest that estimates of the size of the depletion effect should, at the very least, be revised downwards from the effect size reported in bias-uncorrected meta-analyses {Hagger,

2010 #4125}. While the current analysis provides robust evidence that questions the strength of the ego-depletion effect and its replicability, we are not so naïve as to think that the current evidence represents a final nail in the coffin for the ego-depletion effect. We are aware that the current replication provides an important source of data with regards to the effect, but it is but one source. We have outlined possible avenues as to how the research community can move the field forward in providing additional data for the depletion effect and the need to arrive at converging evidence from multiple replication efforts across multiple depletion domains. It is also important to note that the current replication speaks little to the underlying mechanism for the ego-depletion effect, which has also been a subject of considerable debate with proposals for numerous alternative models the limited resource account (Evans, Boggero, & Segerstrom, 2015; Inzlicht & Schmeichel, 2012; Inzlicht, Schmeichel, & Macrae, 2014; Kotabe & Hofmann, 2015). We are also aware of competing literatures such as research on mental fatigue and vigilance (e.g., Gergelyfi, Jacob, Olivier, & Zénon, 2015; Roy, Charbonnier, & Bonnet, 2014) which have yet to be formally unified with the literature on ego-depletion. We call for further co-ordinated research programmes and syntheses that explore the possible mechanisms for the effect and, particularly, moderators that may explain variability in depletion effect sizes across studies and analogs that may assist in mechanistic explanations for the effect.

Acknowledgments

We thank Roy Baumeister, Chandra Sripada, and Daniel Kessler for their assistance in the development of the protocol for the replication studies. We also thank Daniel Kessler for his assistance with the development of the tasks and data analysis. Also, we would like to thank Alex Holcombe for his assistance in developing the protocol and for feedback on a previous version of this paper. Finally, we thank the following people for their help in conducting the experiments: Dayna Bell, T-Jay Anderson, and Kayla Douglas (Birt Lab) [OTHERS TO FOLLOW].

Declaration of Conflicting Interests

The authors declared no conflicts of interest with respect to the authorship or the publication of this article.

Funding

Funding for participant payments and Eprime licenses was provided to individual labs by the Association for Psychological Science via a grant from the Center for Open Science.

References

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology, 74*, 1252-1265. doi: 10.1037/0022-3514.74.5.1252

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

Bush, G., Shin, L. M., Holmes, J., Rosen, B. R., & Vogt, B. A. (2003). The Multi-Source Interference Task: Validation study with fMRI in individual subjects. *Molecular Psychiatry, 8*, 60-70. doi: 10.1038/sj.mp.4001217

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. doi: 10.1037/xge0000083

Carter, E. C., & McCullough, M. E. (2013a). After a pair of self-control-intensive tasks, sucrose swishing improves subsequent working memory performance. *BMC Psychology, 1*, 22. doi: 10.1186/2050-7283-1-22

Carter, E. C., & McCullough, M. E. (2013b). Is ego depletion too incredible? Evidence for the overestimation of the depletion effect. *Behavioral and Brain Sciences, 36*, 683-684. doi: 10.1017/S0140525X13000952

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

Cochran, W. G. (1952). The χ^2 test of goodness of fit. *Annals of Mathematical Statistics, 23*, 315-345. doi: 10.1214/aoms/1177692778

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

Dahm, T., Neshat-Doost, H. T., Golden, A.-M., Horn, E., Hagger, M., & Dalgleish, T. (2011). Age shall not weary us: Deleterious effects of self-regulation depletion are specific to younger adults. *PLoS ONE, 6*, e26351. doi: 10.1371/journal.pone.0026351

Dawson, M. R. W. (1988). Fitting the ex-Gaussian equation to reaction time distributions. *Behavior Research Methods, Instruments, and Computers, 20*, 54-57.

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

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. doi: 10.1086/593291

Gergelyfi, M., Jacob, B., Olivier, E., & Zénon, A. (2015). Dissociation between mental fatigue and motivational state during prolonged mental activity. *Frontiers in Behavioral Neuroscience, 9*. doi: 10.3389/fnbeh.2015.00176

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

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. doi: 10.1037/a0019486

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*, 1539-1558. doi: 10.1002/sim.1186

Inzlicht, M., Gervais, W., & Berkman, E. T. (2015). *News of ego depletion's demise is premature: Commentary on Carter, Kofler, Forster, & McCullough (2015)*. Toronto, Canada: University of Toronto.

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

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

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

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

Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7*, 657-660. doi: 10.1177/1745691612462588

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*. doi: 10.1126/science.aac4716

Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*, 531-536. doi: 10.1177/1745691612463401

Reed, W. R., Florax, R. J. G. M., & Poot, J. (2015). A Monte Carlo analysis of alternative meta-analysis estimators in the presence of publication bias. *Economics Discussion Papers*. 2015-2019.

Roy, R. N., Charbonnier, S., & Bonnet, S. (2014). Detection of mental fatigue using an active BCI inspired signal processing chain. *IFAC Proceedings Volumes (IFAC-PapersOnline), 19*, 2963-2968.

Simonsohn, U. (2009). Publication bias. In M. Borenstein, L. V. Hedges, J. P. T. Higgins & H. R. Rothstein (Eds.), *Introduction to Meta-Analysis*. Chichester, UK: Wiley.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*, 666–681. doi: 10.1177/1745691614553988

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

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

Wan, E. W., & Sternthal, B. (2008). Regulating the effects of depletion through monitoring. *Personality and Social Psychology Bulletin, 34*, 32-46. doi: 10.1177/0146167207306756

Weissman, D. H., Roberts, K. C., Visscher, K. M., & Woldorff, M. G. (2006). The neural bases of momentary lapses in attention. *Nature Neuroscience, 9*, 971-978.

Xu, X., Demos, K. E., Leahey, T. M., Hart, C. N., Trautvetter, J., Coward, P., . . . Wing, R. R. (2014). Failure to replicate depletion of self-control. *PLoS ONE, 9*, e109950. doi: 10.1371/journal.pone.0109950

Table 1

Sample Sizes, Exclusion Information, and Dependent Variable Data for Each Replication of the Ego-Depletion Effect

Lab	Country of participants	Depletion condition								Non-depletion condition							
		Total N	Excluded age	Excluded native language	Excluded other	Excluded accuracy	N included	RTV M (SD)	RT M (SD)	Total N	Excluded age	Excluded native language	Excluded other	Excluded accuracy	N included	RTV M (SD)	RT M (SD)
Birt & Muise	USA	55	3	3	0	22	31	0.31 (0.07)	0.98 (0.14)	55	2	3	0	24	28	0.29 (0.06)	0.94 (0.11)
Brewer & Lau	USA	52	0	0	1	19	47	0.32 (0.08)	0.97 (0.14)	47	0	0	0	13	52	0.31 (0.08)	0.93 (0.13)
Francis & Inzlicht	Canada	23	1	13	0	26	23	0.30 (0.08)	0.86 (0.13)	28	0	12	1	35	27	0.32 (0.09)	0.91 (0.15)
Hagger, Chatzisarantis & Zweinenberg	Australia	71	5	11	0	14	46	0.32 (0.08)	0.93 (0.13)	73	2	8	0	9	55	0.32 (0.08)	0.95 (0.13)
Philipp & Cannon	USA	38	0	0	0	5	38	0.31 (0.08)	0.95 (0.16)	37	0	0	0	6	37	0.31 (0.07)	0.97 (0.13)
Rentzsch, Schutz & Nalis	Germany	62	0	0	0	11	51	0.28 (0.06)	0.88 (0.09)	60	0	1	1	6	52	0.29 (0.06)	0.88 (0.11)
Schlinkert, Schrama & Koole	Netherlands	53	0	0	0	17	36	0.32 (0.08)	0.93 (0.14)	55	0	0	3	9	43	0.32 (0.08)	0.92 (0.12)
Carruth & Miyake	USA	92	0	5	0	32	55	0.32 (0.09)	0.97 (0.14)	93	0	2	0	20	71	0.33 (0.08)	0.97 (0.14)
Ringos	USA	50	1	1	0	17	32	0.33 (0.06)	0.93 (0.05)	50	0	0	1	13	36	0.30 (0.06)	1.00 (0.01)
Crowell, Finley, & Schmeichel	USA	65	0	0	2	29	34	0.32 (0.07)	0.96 (0.14)	65	0	0	1	25	39	0.29 (0.06)	0.96 (0.13)
Evans & Fay	USA	83	1	1	0	41	40	0.32 (0.08)	0.97 (0.15)	84	0	2	0	33	49	0.35 (0.09)	1.03 (0.15)
Calvillo	USA	74	1	7	0	30	36	0.35 (0.08)	1.02 (0.14)	72	0	5	0	28	39	0.32 (0.06)	0.96 (0.15)
vanDellen, Lynch & Campbell	USA	86	0	4	9	31	42	0.34 (0.09)	0.95 (0.15)	86	0	6	12	31	37	0.31 (0.07)	0.88 (0.14)
Cheung, Kroese, Fennis, & de Ridder	Netherlands	102	0	0	1	12	89	0.31 (0.08)	0.96 (0.13)	102	0	0	0	10	92	0.31 (0.07)	0.94 (0.13)
Stamos, Bruyneel & DeWitte	Belgium	59	0	0	0	16	43	0.30 (0.07)	0.94 (0.15)	58	2	0	0	6	50	0.31 (0.07)	0.94 (0.14)
Brandt	Netherlands	85	0	10	0	32	48	0.28 (0.07)	0.88 (0.13)	80	2	7	0	20	54	0.29 (0.07)	0.89 (0.13)
Lange, Heise & Hoemann	Germany	60	0	0	0	6	54	0.28 (0.06)	0.88 (0.10)	60	0	0	0	8	52	0.30 (0.07)	0.90 (0.15)
Ullrich & Primoceri	Switzerland	59	0	0	2	7	50	0.29 (0.06)	0.90 (0.12)	62	0	3	1	5	53	0.29 (0.06)	0.89 (0.11)
Yusainy, Wimbarti, Nurwanti & Octavianus	Indonesia	100	0	0	1	17	82	0.29 (0.08)	0.91 (0.13)	100	0	0	1	25	74	0.27 (0.06)	0.89 (0.13)
Zerhouni & Muller	France	56	0	0	1	23	32	0.34 (0.07)	1.04 (0.13)	57	0	1	0	10	46	0.38 (0.08)	1.09 (0.14)
Elson	Germany	52	2	0	0	9	42	0.30 (0.07)	0.96 (0.14)	54	0	0	1	5	48	0.30 (0.07)	0.95 (0.15)
Otgaar, Martijn, Alberts, Howe & Michirev	Netherlands	50	0	0	2	25	25	0.28 (0.07)	0.86 (0.12)	50	0	0		6	44	0.31 (0.07)	0.93 (0.14)
Wolff, Muzzi & Brand	Germany	55	0	0	1	16	38	0.33 (0.07)	0.97 (0.12)	56	0	0	1	6	49	0.3 (0.06)	0.92 (0.12)
Tinghög & Koppel	Sweden[a]	51	2	0	0	10	40	0.27 (0.05)	0.86 (0.11)	51	0	0	0	9	42	0.29 (0.06)	0.92 (0.13)

Note. Exclusion columns are not mutually exclusive e.g., some participants were excluded because they failed to meet age and language criteria. RTV = Reaction time variability on incongruent items of the multi-source interference task (MSIT), RT = Overall reaction time on incongruent items on the MSIT. [a]Conducted on fluent English-speaking students in Sweden.

Table 2

Results of Meta-Analysis of Replications of Ego-Depletion Effect

Dependent variable	*d*	CI_{95}		SE	*Q*	*p*	I^2	τ
		LL	UL					
RTV	0.020	-0.089	0.130	.022	38.111	.025	39.651	0.170
RT	-0.063	-0.215	0.088	.077	72.724	<.001	68.374	0.310
Letter 'e' accuracy	-1.814	-1.963	-1.665	.076	50.715	.001	54.648	0.272
Self-report measures								
Effort	0.779	0.630	0.929	.076	66.169	<.001	65.240	0.299
Fatigue	0.090	-0.018	0.197	.055	36.906	.033	37.679	0.163
Difficulty	1.919	1.714	2.123	.104	92.415	<.001	75.112	0.438
Frustration	0.815	0.665	0.966	.077	66.674	<.001	65.504	0.301

Note. In all cases number of studies was 24. RTV = Reaction time variability on incongruent items of the multi-source interference task (MSIT), RT = Overall reaction time on incongruent items on the MSIT. *d* = averaged corrected standardized mean difference across ego-depletion and control groups; CI_{95} = 95% confidence intervals of *d*; LL = Lower limit of confidence interval; UL = Upper limit of confidence interval; SE = Standard error of *d*; *Q* = Cochran's (1952) *Q* Statistic; I^2 = Higgins and Thompson's (2002) I^2 statistic; τ = Tau statistic.

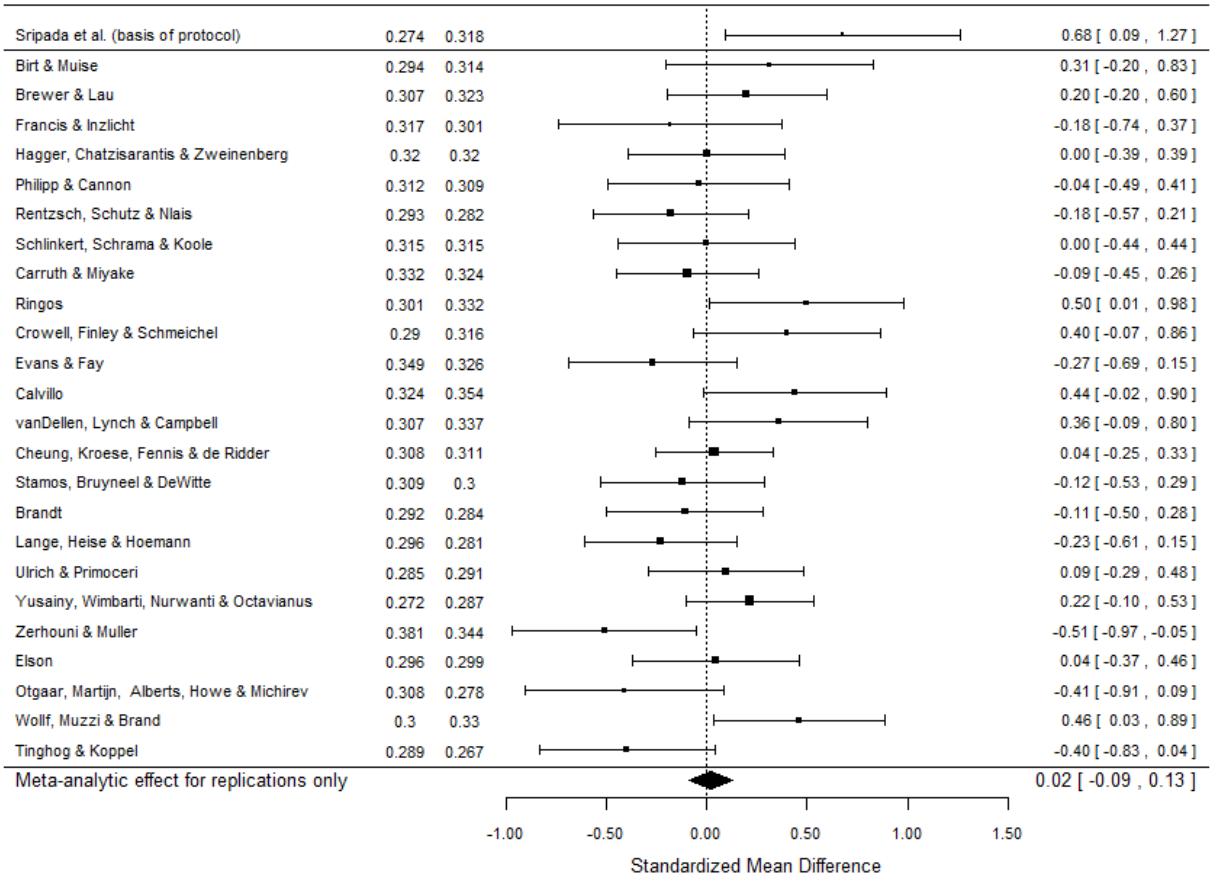

Figure 1. Forest plot of the effect of depletion condition on RTV (reaction time variability) for the multi-source interference task with positive effects indicating greater depletion. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control groups and a forest plot of the standardized mean difference scores with 95% confidence intervals. The overall meta-analytic effect size does not include Sripada et al. (2014).

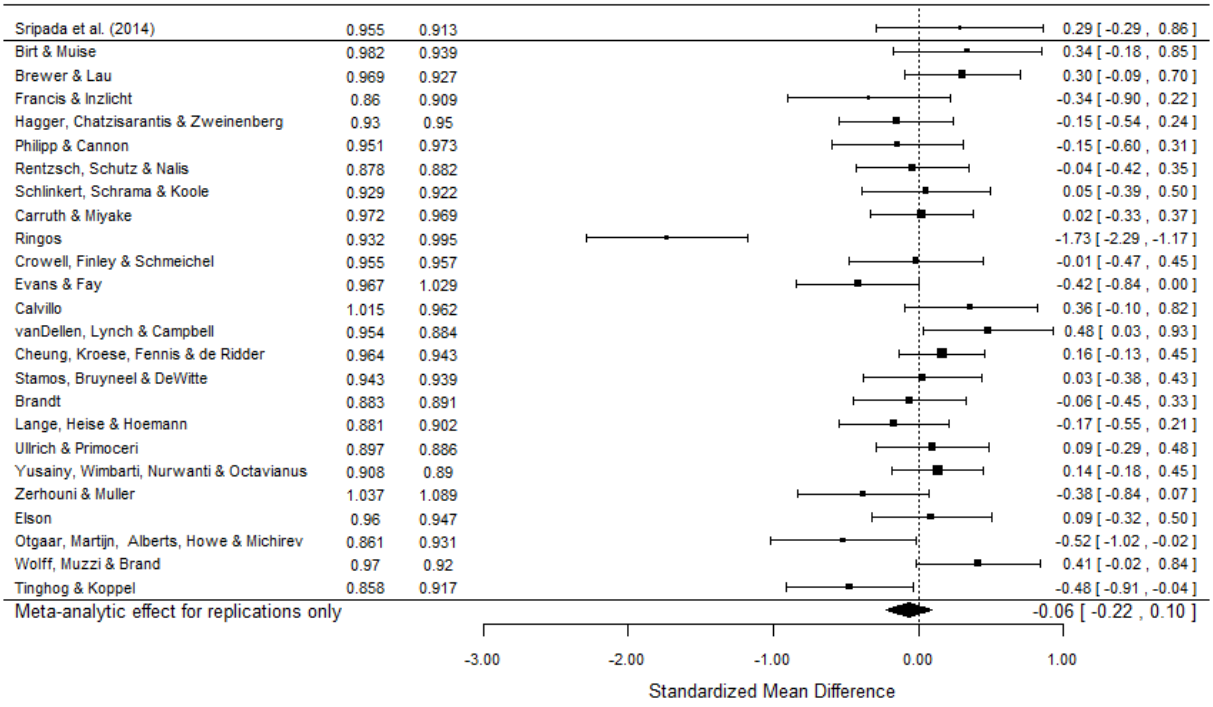

Figure 2. Forest plot of the effect of depletion condition on reaction time (RT) for the multi-source interference task with positive effects indicating greater depletion. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control groups and a forest plot of the standardized mean difference scores with 95% confidence intervals. The overall meta-analytic effect size does not include Sripada et al. (2014).

Appendix A. Individual lab details.

Martin S. Hagger, Curtin University, Australia
Nikos L. D. Chatzisarantis, Curtin University, Australia
Maria Zweinenberg, Curtin University, Australia and University of Bordeaux, France
OSF: https://osf.io/quwx9/

Participants (N = 144, males = 28, females = 116, *M* age = 20.83 years, *SD* = 4.55) were recruited from an undergraduate participant pool at Curtin University. All participants were enrolled in an undergraduate degree programme in psychology and participated in the study for course credit. Twenty-six participants were excluded from the final analysis because they did not meet the meet the specified inclusion criteria for first-spoken language (n = 19) and age (n = 7). A further 17 participants were excluded because their performance on the letter 'e' task and MSIT fell below 80% accuracy or had mean reaction time or mean reaction time variability values that fell outside two standard deviations of the sample mean on the MSIT. The final sample (N = 101, males = 21, females = 80, *M* age = 20.01 years, *SD* = 2.36) comprised 46 participants in the hard letter 'e' (ego-depletion) condition and 55 participants in the easy letter 'e' (control) condition. Maria Zweinenberg, Nikos L. D. Chatzisarantis, and Martin S. Hagger served as the experimenters. Zweinenberg was naïve to the purpose of the experiment and ran the majority of the participants overall (N = 111) and the participants included in the final analysis (N = 78). None of the experimenters were blinded to condition assignment because they read out instructions to the participants. Our procedures followed the approved protocol and did not deviate from our preregistered plan.

Appendix B. Results of meta-analysis of letter ‘e’ task performance and self-report ratings of effort, fatigue, difficulty, and frustration.

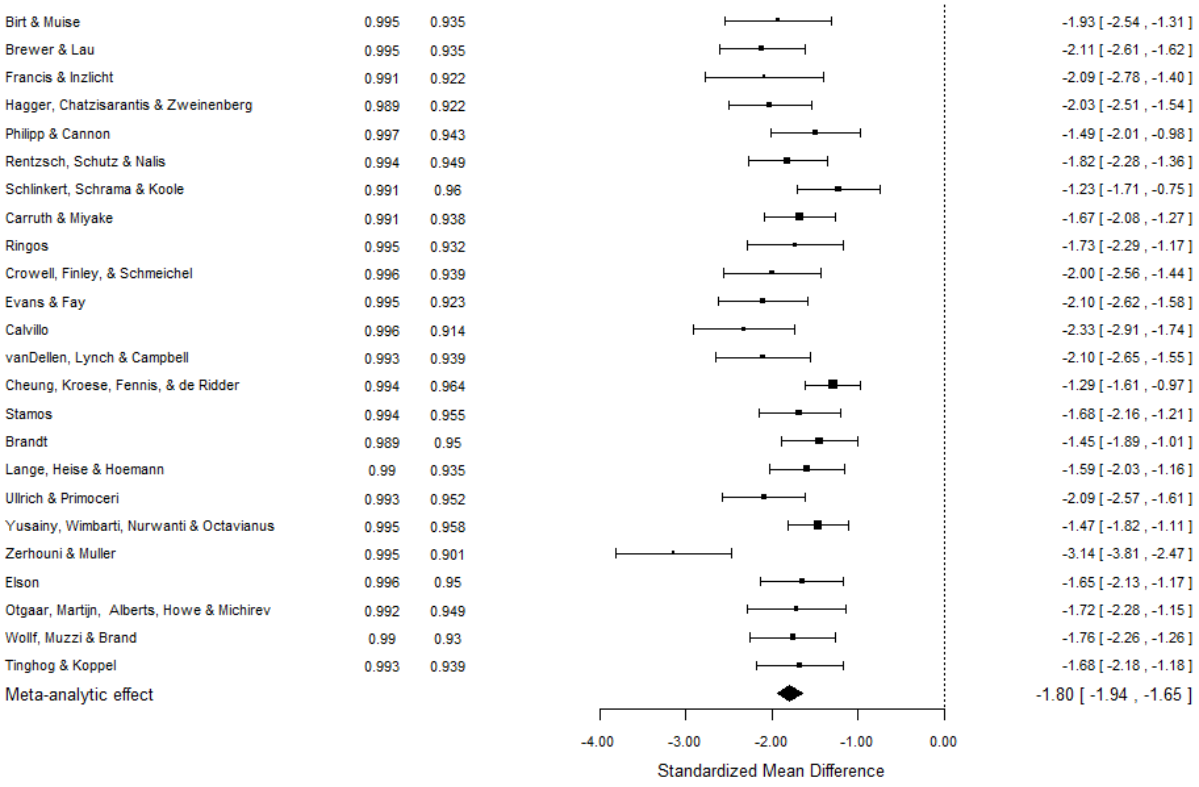

Figure B1. Forest plot of the effect of depletion condition on accuracy rates on the letter ‘e’ task with positive effects indicating greater accuracy. For each lab, the figure shows the standardized mean difference (Cohen’s *d*) across depletion and control groups and a forest plot of the standardized mean difference scores with 95% confidence intervals.

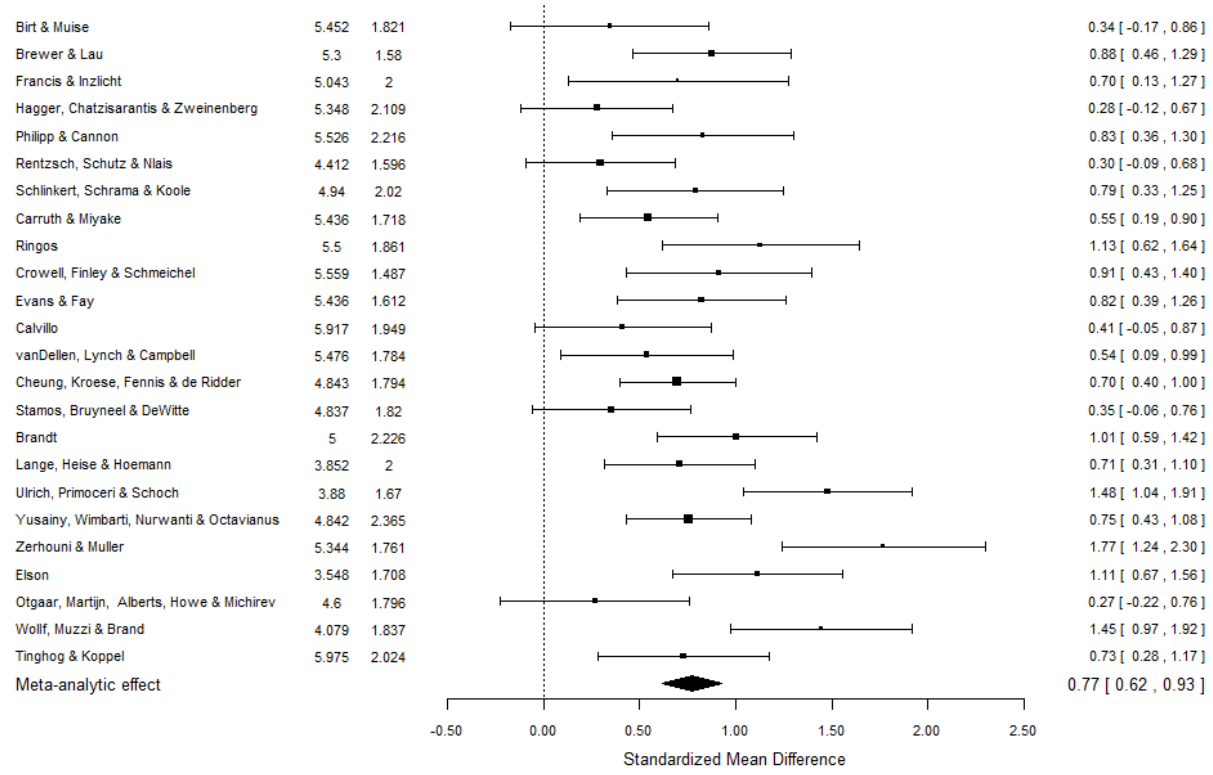

Figure B2. Forest plot of the effect of depletion condition on self-reported effort. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control groups and a forest plot of the standardized mean difference scores with 95% confidence intervals.

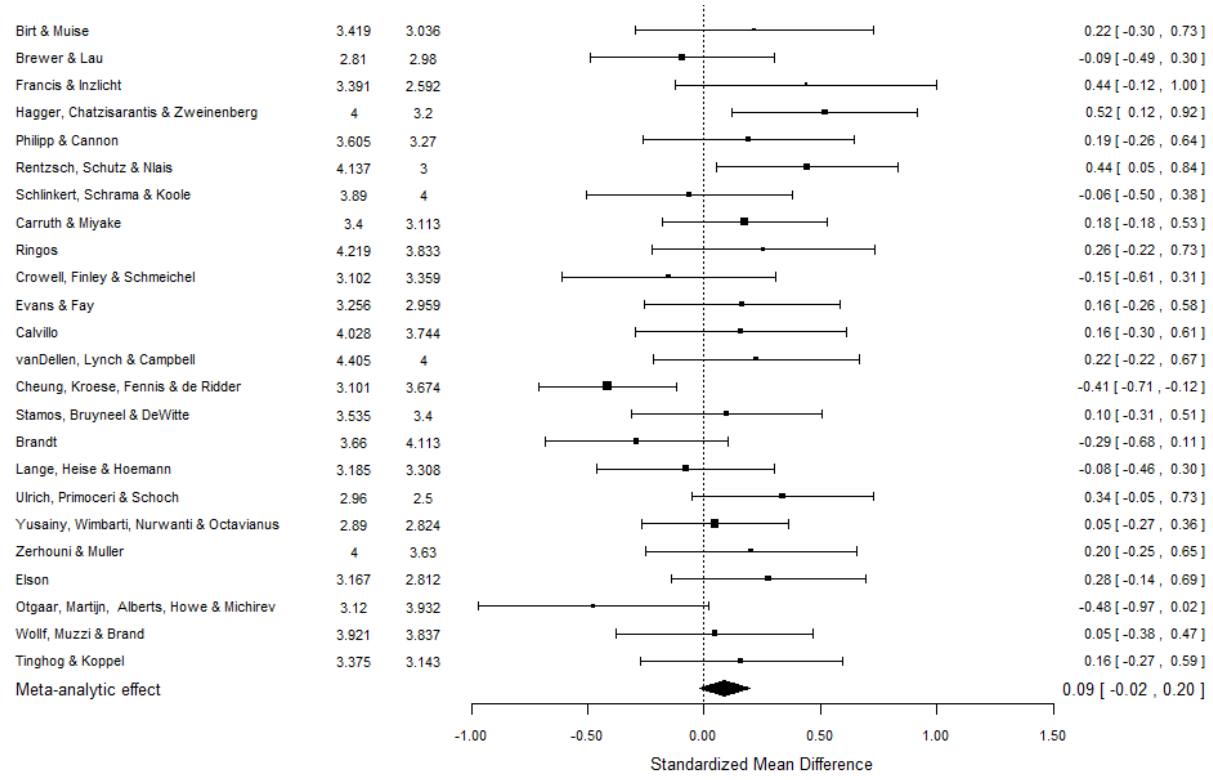

Figure B3. Forest plot of the effect of depletion condition on self-reported fatigue. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control groups and a forest plot of the standardized mean difference scores with 95% confidence intervals.

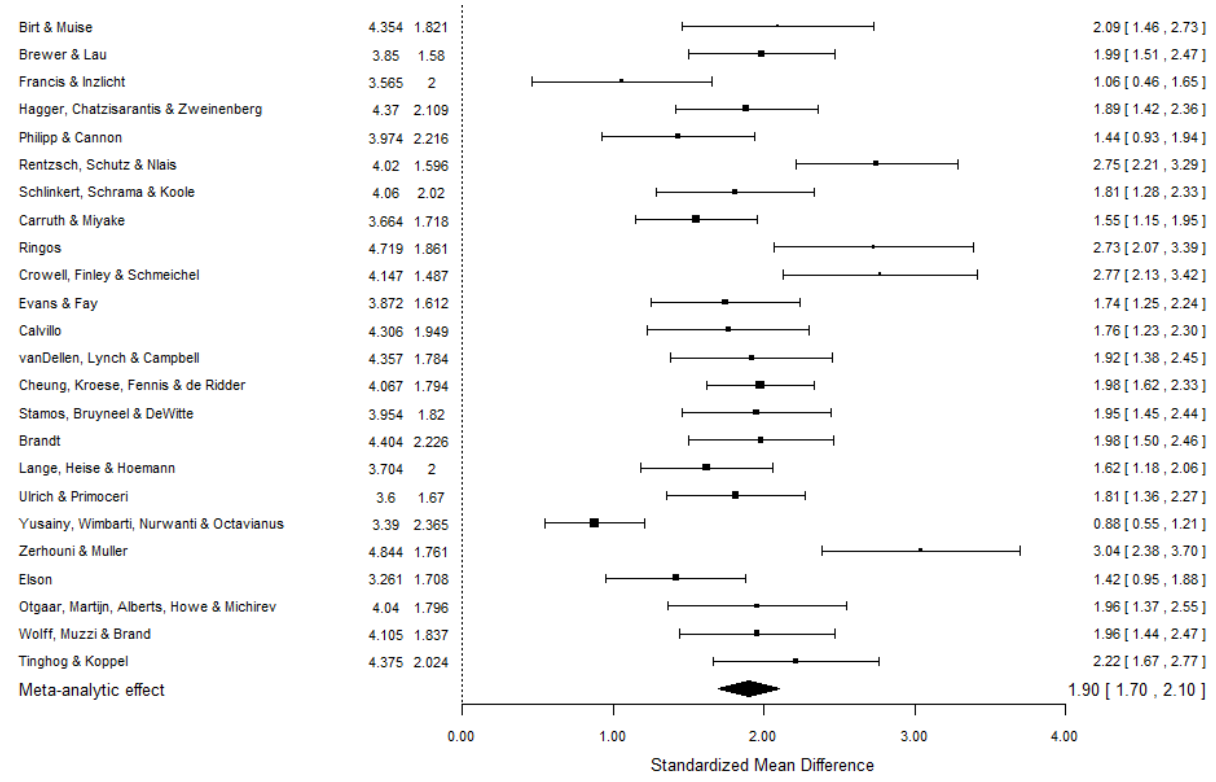

Figure B4. Forest plot of the effect of depletion condition on self-reported difficulty. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control groups and a forest plot of the standardized mean difference scores with 95% confidence intervals

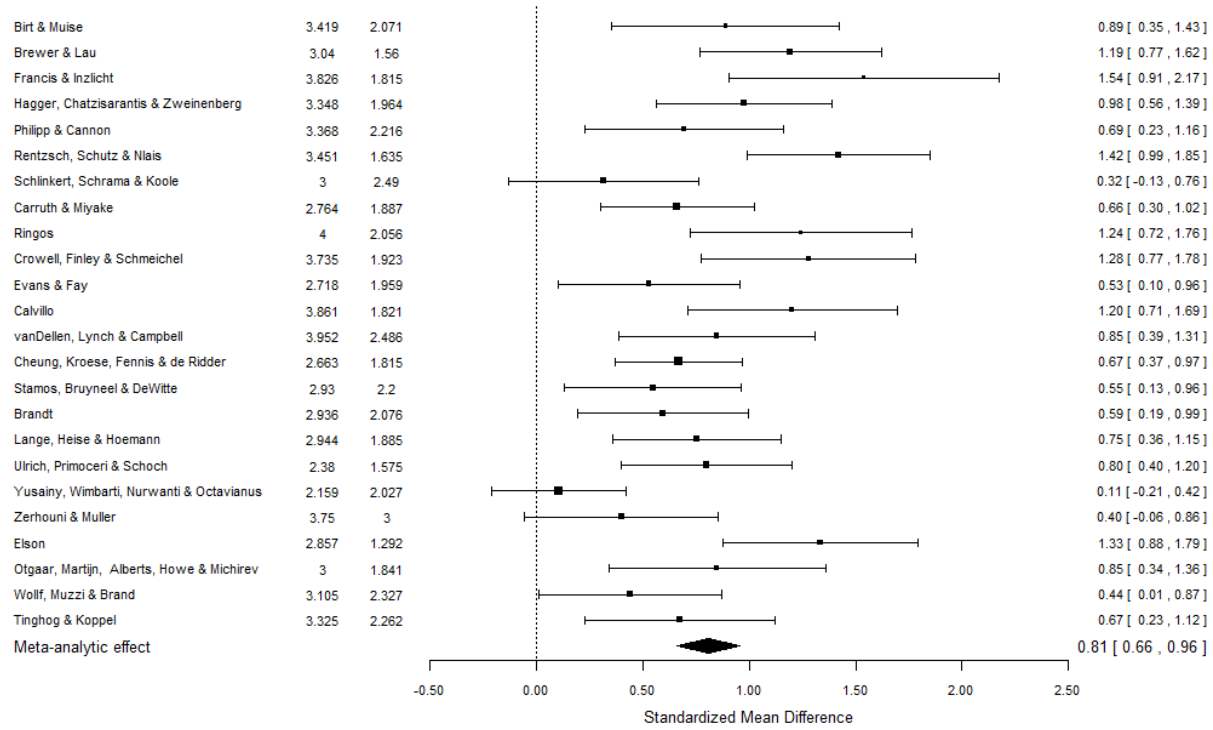

Figure B5. Forest plot of the effect of depletion condition on self-reported frustration. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control groups and a forest plot of the standardized mean difference scores with 95% confidence intervals.