A multilab preregistered replication of the ego-depletion effect

M. S. Hagger,* N. L. D. Chatzisarantis,* H. Alberts, C. O. Anggono,

C.Batailler, A. R. Birt, R. Brand, M. J. Brandt, G. Brewer, S. Bruyneel,

D.P. Calvillo, W. K. Campbell, P. R. Cannon, M. Carlucci, N. P. Carruth,

T. Cheung, A. Crowell, D. T. D. De Ridder, S. Dewitte, M. Elson, J. R. Evans,

B. A. Fay, B. M. Fennis, A. Finley, Z. Francis, E. Heise, H. Hoemann,

M. Inzlicht, S. L. Koole, L. Koppel, F. Kroese, F. Lange, K. Lau, B. P. Lynch,

C. Martijn, H. Merckelbach, N. V. Mills, A. Michirev, A. Miyake, A. E. Mosser,

M. Muise, D. Muller, M. Muzi, D. Nalis, R. Nurwanti, H. Otgaar, M. C. Philipp,

P. Primoceri, K. Rentzsch, L. Ringos, C. Schlinkert, B. J. Schmeichel, S. F. Schoch, M. Schrama, A. Schütz, A. Stamos, G. Tinghög, J. Ullrich, M. vanDellen,

S. Wimbarti, W. Wolff, C. Yusainy, O. Zerhouni, and M. Zwienenberg
*) Proposing authors

Journal Article

N.B.: When citing this work, cite the original article.

Original Publication:

Martin S. Hagger, Nikos L. D. Chatzisarantis, et al: A multilab preregistered replication of the ego-depletion effect, Perspectives on Psychological Science, 2016. 11(4), pp.546-573.
http://dx.doi.org/10.1177/1745691616652873
Copyright: Association for Psychological Science
http://www.psychologicalscience.org/

Postprint available at: Linköping University Electronic Press

http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-132383

LINKÖPINGS UNIVERSITET

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, Cédric Batailler, Angela Birt, Ralf Brand, Mark J. Brandt, Gene Brewer, Sabrina Bruyneel, Dustin P. Calvillo, W. Keith Campbell, Peter R. Cannon, Marianna Carlucci, Nicholas Carruth, Tracy Cheung, Adrienne Crowell, Denise T. D. De Ridder, Siegfried Dewitte, Malte Elson, Jaqueline R. 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, Harald Merckelbach, Nicole V. Mills, Alexej Michirev, Akira Miyake, Alexandra E. Mosser, Megan Muise, Dominique Muller, Milena Muzi, Dario Nalis, Ratri Nurwanti, Henry Otgaar, Michael Philipp, Pierpaolo Primoceri, Katrin Rentzsch, Lara Ringos, Caroline Schlinkert, Brandon J. Schmeichel, Sarah F. Schoch, Michel Schrama, Astrid Schütz, Angelos Stamos, Gustav Tinghög, Johannes Ullrich, Michelle vanDellen, Supra Wimbarti, Wanja Wolff, Cleoputri Yusainy, Oulmann Zerhouni, Maria Zwienenberg

Protocol vetted by: Chandra Sripada, Daniel Kessler, Roy Baumeister
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., Batailler, C., Birt, A., … Zwienenberg, M. (2015). A multi-lab pre-registered replication of the ego-depletion effect. *Perspectives on Psychological Science,* X, X-X.

Unedited Manuscript

Abstract

Good self-control has been linked to adaptive outcomes such as better health, cohesive personal relationships, success in the workplace and at school, and less susceptibility to crime and addictions. In contrast, self-control failure is linked to maladaptive outcomes. Understanding the mechanisms by which self-control predicts behavior may assist in promoting better regulation and outcomes. A popular approach to understanding self-control is the strength or 'resource depletion' model. Self-control is conceptualized as a limited resource which becomes depleted after a period of exertion resulting self-control failure. The model has typically been tested using a 'sequential-task' experimental paradigm in which people completing an initial self-control task have reduced self-control capacity and poorer performance on a subsequent task, a state known as 'ego depletion'. Although a meta-analysis of ego-depletion experiments found a medium-sized effect, subsequent meta-analyses have questioned the size and existence of the effect and identified instances of possible bias. The analyses served as a catalyst for the current registered replication report of the ego-depletion effect. Multiple laboratories (k = 23, total N = 2141) conducted replications of a standardized ego-depletion protocol based on a sequential-task paradigm by Sripada et al. Meta-analysis of the studies revealed that the size of the ego-depletion effect was small with 95% confidence intervals that encompassed zero. We discuss implications of the findings for the ego-depletion effect and the resource depletion model of self-control.

Unedited Manuscript

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

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

Good self-control is important for optimal human functioning. Self-control has been regarded as an individual's capacity to actively override or inhibit impulses, suppress urges, resist temptations, and break ingrained, well-learned behaviors, or habits. Self-control therefore reflects the extent to which an individual can override a dominant response in favor of an alternative, more effortful course of action. Good self-control has been linked to adaptive outcomes in multiple domains including school, the workplace, social relationships, and health (de Ridder, Lensvelt-Mulders, Finkenauer, Stok, & Baumeister, 2012; Dvorak & Simons, 2009; Hagger, Wood, Stiff, & Chatzisarantis, 2010b; Tangney, Baumeister, & Boone, 2004). Analogously, poor self-control is associated with many maladaptive outcomes including poor health, financial instability, dysfunctional social relationships, and susceptibility to drug abuse and crime (Baumeister & Alquist, 2009; Baumeister, Heatherton, & Tice, 1994; Wills, Isasi, Mendoza, & Ainette, 2007). Accordingly, it is vital to understand why people may succeed or fail at self-control.

The conceptualization that self-control capacity depends on a finite resource has gained considerable attention in the literature. 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 'sequential-task' experimental paradigm in which participants engaged in two consecutive tasks. For participants randomly allocated to the experimental (ego depletion) group, both tasks both required self-control. For participants

Unedited Manuscript

allocated to the control (no depletion) group only the second task required self-control while the first task did not require any, or very little, self-control. The self-control tasks used required participants to alter or modify an instinctive, well-learned response, akin to resisting an impulse or temptation (Baumeister, Vohs, & Tice, 2007).

Consistent with the predictions of the resource depletion model, 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 conceptual replications of the original findings using the sequential-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, 2010a).

However, recent conceptual and empirical analyses have challenged the resource depletion explanation for the self-regulatory failures observed in ego-depletion experiments and questioned the strength of the ego-depletion effect or whether it exists at all. Recent analyses have suggested that original meta-analytic effect size for ego depletion may be inflated. Re-analyses of Hagger et al.'s meta-analytic findings (Carter & McCullough, 2013b; Carter & McCullough, 2014) and a new meta-analysis of tests of the ego-depletion effect that included 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). These regression techniques have been proposed as means to detect bias in sets of studies included in meta-analyses, known as 'small study' bias. Small study bias refers to increased likelihood of improbably high effect sizes relative to study precision in a sample of studies included in a meta-analyses. The bias may be indicative of publication bias, that is, the propensity of journal editors to favour publication of

Unedited Manuscript

studies that achieve statistical significance and tend to have larger effect sizes relative to their sample size (Sterne, Egger, & Davey Smith, 2001).

Carter et al.'s analyses revealed substantial 'small study' bias in the effect size reported in Hagger et al.'s (2010a) original meta-analysis and indicated that many published studies included in the original analysis, and in their updated meta-analysis, were substantially underpowered suggesting that the likelihood of finding of so many large, statistically significant effects was improbable. In both their re-analysis and updated meta-analysis, Carter et al. (2015) suggested that, based on their regression analyses, 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 indicated that a facilitation effect may occur in which task performance improves with prior self-control in multi-task experiments (e.g., Converse & DeShon, 2009; Dewitte, Bruyneel, & Geyskens, 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 interpretation of the regression techniques was misleading in the presence of substantial heterogeneity in the effect size. This might be the case if, for example, the true effect is larger in smaller studies (Sterne et al., 2001). Furthermore, the regression techniques are based on the 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., where there is

Unedited Manuscript

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 indicate the existence of bias in meta-analytically derived effect sizes attributable to small study effects, they cannot definitively identify the source of the bias (Simonsohn, 2009).

Issues of interpretation notwithstanding, the existence 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 sequential-task paradigm, as advocated by Carter and McCullough (2013b; 2014) and Hagger and Chatzisarantis (2014).

Protocol Development

While the sequential-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 tasks used is that there is no single agreed standardized set of tasks for use in sequential-task paradigm tests of the ego-depletion effect.

A further issue in developing the protocol was the need for tasks to be sufficiently standardized to rule out, wherever possible, idiosyncratic lab-specific differences in the presentation of tasks or other variations that may reduce the consistency of the protocol

Unedited Manuscript

implementation across labs. Whereas typical practice in registered replications of psychological research has tended to prioritize the replication of the original experiment (e.g., Alogna et al., 2014; Eerland et al., 2016), the tasks used in the original experiments were deemed too elaborate or complex to be appropriate for a multi-lab replication. For example, one of the tasks used to deplete self-control resources in the original tests of the ego-depletion effect required participants to taste radishes and resist cookies (Baumeister et al., 1998, Study 1). This task would require extensive experimenter involvement in its delivery which may increase variability across labs. Similarly, persistence on unsolvable anagrams (Baumeister et al., 1998, Study 3) is likely to 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 sequential-task procedure that adopted standardized 'behavioral' tasks requiring little adaptation across labs and minimal interpersonal involvement by the experimenter.

Given these concerns, we sought to identify a previously-published procedure that was in keeping with original sequential-task tests of the ego-depletion effect, but could be standardized for a multi-lab replication so as to minimize experimenter input and methodological variability across laboratories. The ego-depletion paradigm adopted by Sripada, Kessler, and Jonides (2014) was identified as one that fit well with our requirements: 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 to use these tasks was based on the recommendation of Roy Baumeister. The protocol was developed in close consultation with Chandra Sripada and Daniel Kessler, co-authors of the original experiment, who made the tasks and procedure used in the original study available for the replication project. It is important to note that Sripada et

Unedited Manuscript

al.'s original study also examined the effects of the 'study drug' Ritalin (methylphenidate) on ego-depletion in a 2 x 2 placebo-controlled experimental design. 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 in the context of their depletion paradigm. These authors found a statistically significant effect for ego-depletion (d = 0.69).

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, 2014. 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 data collection due to technical difficulties or insufficient resources (e.g., access to participants or research assistants) leaving 24 labs contributing to the project. Participating labs pre-registered their implementation plan on the Open Science Framework and conducted independent replications. Each implementation plan was vetted by the registered replication reports editor (Alex Holcombe) for consistency with the protocol prior to data collection. Participating labs were in Australia, Belgium, Canada, France, Germany, Indonesia, the Netherlands, New Zealand, Sweden, Switzerland, and the United States. Co-ordinated and systematic translation efforts were undertaken to prepare study materials in labs recruiting participants whose native language was not English. The investigators of each participating lab had 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.

Unedited Manuscript

Protocol Requirements

In this section we provide details of the replication protocol. From the 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 undergraduate students who participated in return for course credit or payment. Participants were recruited from institution-managed participant pools or in response to study advertisements. Based on a statistical power analysis with alpha at 0.01 and beta at 0.95, we computed that a sample size of 168 participants, with 84 in each of the depletion and non-depletion conditions, was required to detect the medium effect size (Hagger et al., 2010a). While we strongly recommended that participating laboratories' replications met this sample size, replications with sample sizes of 100 participants was considered the guideline minimum (≥ 50 participants in each condition). Most labs were able to achieve this target in their recruitment, but due to the rigorous exclusion criteria for the tasks used in the sequential-task paradigm, the targeted sample size was not achieved in some cases. Given evidence suggesting that older participants show a weaker ego-depletion effect (Dahm et al., 2011), participants were required to be between 18 and 30 years old. As study materials were language specific, participants were required to be native speakers of the language in which the replication was conducted. Participants from labs in English-speaking countries (Australia, Canada, New Zealand, and the United States) were excluded if they did not report English as their first and primary language. Labs in non-English speaking countries conducted the replication using study materials translated into the primary language of the participants and non-native speakers were excluded. One lab conducted the replication in a sample of English-speaking students in Sweden (Tinghög & Koppel). While the participants from this lab were fluent English

Unedited Manuscript

speakers, their results were omitted from the final analysis because they deviated from the native language inclusion criterion leaving 23 labs included in the final analysis[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 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) ego-depletion paradigm. 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, it was recommended that labs randomized participants 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.

[1]Supplementary analyses that include data from the Tinghög and Koppel lab can be found on the replication OSF site: https://osf.io/4zy8k/

Unedited Manuscript

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, participants completed self-report items measuring effort, fatigue, difficulty, and frustration on the first task, which were used as manipulation checks for the ego-depletion manipulation. Participants then completed the second task. In an exit questionnaire, participants' thoughts on the purpose of the experiment were probed. They were then thanked and informed that the experiment had concluded. Some of the participating labs' IRBs 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. Others provided a debrief immediately after the experiment but asked participants not to share details with fellow students.

Overall, there were two differences between the current replication protocol and the original protocol by Sripada et al. (2014): (1) we did not administer a capsule prior to the task

Unedited Manuscript

protocol, where participants expected it to contain either placebo or Ritalin, and (2) we administered self-report measures of task effort, fatigue, difficulty, and frustration after the first task, while no measures were administered in the original study. The self-report measures were included to check whether the initial task was subjectively arduous and depleting for participants assigned to the ego-depletion group relative to the control group. Similar measures such as these have been administered in many ego-depletion experiments, including the original ego-depletion experiments (Baumeister et al., 1998; Muraven et al., 1998).

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* (available at https://osf.io/ifdj3/). Two versions were used: depletion and no depletion. In the depletion version, 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' was displayed and withhold the response if the 'e' was next to or one letter away from a vowel. The no depletion version was matched in all respects with the exception that participants were required to press a button whenever a word with the letter 'e' was 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 20 practice trials before the commencement of the experiment. The main session comprised 150 trials and lasted 7 minutes and 30 seconds. Participants' reaction times and errors for the letter 'e' task were recorded. The depletion version of the letter 'e' task was considered to be more demanding, and require greater self-control, than the no depletion version because participants had to inhibit the tendency to respond to any 'e' and instead apply the more restrictive rules.

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

Unedited Manuscript

https://osf.io/ifdj3/). Numeric stimuli were presented on the computer screen with participants making responses using the keyboard. The stimuli were sets of three digits comprising combinations of the numerals 1, 2, 3 or 0. 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 would be presented with sets of three digits 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' or 'congruent' 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 congruent 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 was also caused by varying the size of the digits in the set. In the congruent version variation in the digit size was always consistent with the target digit, for example the target digit was always the larger or smaller digit relative to the other digits in the set. In the incongruent version, variation in digit size was not always consistent with the target digit, requiring the participant to inhibit both the position and size of the target digit in favor of its identity. Participants were asked to respond as quickly and accurately as possible. Participants completed 20 practice MSIT trials before the commencement of the experiment. The main 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.

Unedited Manuscript

Performance on the MSIT comprised the dependent measure of self-control. The MSIT provides two measures of performance: 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 current protocol. RTV is considered an analog of attentional control. Participants with good attentional control are effective in maintaining task-directed focus and supressing 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 incongruent trials on the MSIT (Weissman, Roberts, Visscher, & Woldorff, 2006). While this should also inflate mean RT, RTV 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 bilingual translator followed by back-translation by an independent fluent bilingual translator. The translated versions were also independently reviewed by the replication proposer (Martin Hagger) and registered replication reports editor (Alex Holcombe). The specific translation procedures of each non-English speaking lab are documented on their respective OSF webpages. 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.

Unedited Manuscript

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 met protocol requirements. Participant exclusion criteria were specified prior to 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, or failed 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 excluded from the analysis are provided on participating labs' OSF webpages (https://osf.io/jymhe/).

Critical comparisons

By convention in sequential-task paradigm studies examining the ego-depletion effect, the critical analysis is a one-way test of difference on task performance across the depletion and no-depletion groups. In the current replication, the primary dependent variable was RTV for incongruent trials of the MSIT and the critical test was whether RTV was higher for participants assigned to the depletion condition relative to those assigned to no depletion condition. This is identical to the critical test 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 (k = 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 (k = 10) indicated it would be

Unedited Manuscript

smaller than reported in previous analyses, but greater than $d = 0.15$. One lab predicted a null effect[2].

Additional analyses were planned on the secondary dependent variable and the control (manipulation check) variables: mean RT for incongruent items on the MSIT, mean RT for 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. It is important to note that Sripada et al. found no effect on RT and considered the RTV a better indicator of self-control failure as it was hypothesized to closely reflect levels of attentional control. Poorer accuracy and greater levels of effort, difficulty, fatigue and frustration in the depletion version of the letter 'e' task condition relative to the no depletion version would indicate that participants found the depletion version more arduous and effortful.

Unedited Manuscript

Results

Lab Demographics and Preliminary Analyses

Sample demographics and results for each of the participating labs ($k = 23$, total $N = 2141$) are provided in Table 1 for the depletion and non-depletion conditions alongside the ego-depletion data from the replicated study for comparison (Sripada et al., 2014). The table provides sample sizes, details of exclusions and reasons, and the means and standard deviations of the mean RTV and mean RT dependent variables in each condition. Demographic details of participants and reasons for exclusion, experimenters' details, and deviations from preregistered protocol for all participating labs can be found in Appendix A. Analysis of rates of exclusions for inaccuracy on letter 'e' and MSIT tasks revealed significant differences in the proportion of participants excluded for low accuracy (<80% accuracy on tasks) relative to

[2]Full details of the expectations and experience of all participating labs can be found on the replication OSF site: https://osf.io/atxbi/

inclusions across depletion and no-depletion groups in five of the 23 laboratories. These data suggest that rates of exclusion due to accuracy were largely independent of condition allocation. Details of these supplementary analyses are provided in Appendix B.

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 a 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 dependent and control variables across conditions in pooled data from the labs were tested using separate meta-analyses. 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 statistically 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 of the target dependent variables (mean RTV and RT for the MSIT) 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 are provided in Figures 1 (RTV) and 2 (RT) alongside effect-size data for the placebo condition of the

Unedited Manuscript

Sripada et al. study for comparison. Positive effect sizes for RT and RTV represent the extent of a relative deficit in performance on the second task in the depletion group and thus 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 C. Summary statistics from the meta-analyses for all dependent variables are presented in Table 2[3].

Averaged sample-weighted effect sizes for the mean RTV (d = 0.04, 95% confidence interval: -0.07 to 0.15) and RT (d = 0.04, 95% confidence interval: -0.07 to 0.14) variables were small and confidence intervals included the value of zero. In terms of individual labs' data, only three of the 23 replications did not have 95% confidence intervals for the ego-depletion effect size that included zero for RTV, and one of those was negative (i.e., in the opposite direction to the hypothesized ego-depletion effect). Similarly, only three labs found mean RT values with confidence intervals that did not include the value of zero, two of which were negative. We also found moderate levels of heterogeneity in the effect sizes for mean RTV (I^2 = 36.08%, Q_{22} = 33.42, p = .045) and RT (I^2 = 34.13%, Q_{22} = 33.40, p = .056) indicating substantial variability in the effect across labs after correction for methodological artifacts (i.e., sampling error). This finding suggests the presence of other extraneous variables that may moderate the effect size across laboratories, despite all labs running the experiment with 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.

Unedited Manuscript

[3]The stringent inclusion criteria based on accuracy rates on the letter 'e' and MSIT tasks resulted in relatively high rates of participant exclusion across labs. A possible concern with the high exclusion rates is that participants with low accuracy on tasks were more vulnerable to depletion, which may have masked the effect. Given the potential for the exclusion rates to affect the outcome, we conducted post hoc analyses identical to the planned analyses in which participants previously excluded for accuracy were included. The analyses revealed very similar results to the analyses including the exclusions with small close-to-zero effects for RTV and RT. Full analyses are reported in Appendix B and the data and analysis files are provided on the OSF website under supplementary analyses: https://osf.io/4zy8k/

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. This moderator analysis tested the hypothesis that the use of translated versions of the letter 'e' task introduced method variance to the ego-depletion effect. Results of the separate meta-analyses for the English and non-English speaking labs are provided in Table 2. While there were only very small observed differences in effect sizes in the English speaking labs' data for the mean RTV (d = 0.14, 95% confidence interval: -0.02 to 0.30) and RT (d = 0.08, 95% confidence interval: -0.09 to 0.24) dependent variables relative to the non-English speaking labs (RTV: d = -0.04, 95% confidence interval: -0.18 to 0.10; RT: d = 0.002, 95% confidence interval: -0.14 to 0.15), the moderator analysis served to produce homogenous cases in both the English speaking (I^2 = 30.45%, Q_{10} = 14.38, p = .156) and non-English speaking (I^2 = 34.82%, Q_{11} = 16.88, p = .112) labs for mean RTV. The analysis also produced a homogenous case for the non-English speaking labs (I^2 = 20.38%, Q_{11} = 13.82, p = .243), but not the English speaking labs (I^2 = 47.84%, Q_{10} = 19.17, p = .038), for RT. The homogenous effect sizes were based on the Q-statistic suggesting that the variability in the effect sizes attributable to methodological artifacts (i.e., sampling error) was no different to the overall variability in the effect size across samples. However, it is important to note that the I^2 statistic, often considered a better indicator of heterogeniety (Higgins & Thompson, 2002), indicated moderate heterogeniety in the effect sizes. Importantly, there was substantial overlap in the confidence intervals of each moderator group and all encompassed zero as a possible value.

Forest plots for the meta-analyses of participants' accuracy on the letter 'e' task and self-report ratings of effort, fatigue, difficulty, and frustration are presented in Appendix C (see Figures C1-C5). We found large effects for the depletion condition on letter 'e' task accuracy (d = -1.82, 95% confidence interval: -1.98 to -1.67), and scores on effort (d = 0.78, 95% confidence interval: 0.63 to 0.94), difficulty (d = 1.91, 95% confidence interval: 1.70 to 2.12),

Unedited Manuscript

and frustration (d = 0.82, 95% confidence interval: 0.67 to 0.98), but a substantially smaller effect for fatigue with confidence intervals that included zero (d = 0.09, 95% confidence interval: -0.03 to 0.20). Overall, these findings provide some evidence that the depletion version of the letter 'e' task was more effortful and aversive than the no depletion version.

Discussion

The current report presents the first registered multi-lab replication of the ego-depletion effect. Results across 23 (N = 2141) participating 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 smaller than the ego-depletion effect size for RTV in the placebo condition of the Sripada et al.'s (2014) study (d = 0.69), that the present protocol was based on. The present effects are also much smaller than the uncorrected ego-depletion effect sizes reported in Hagger et al.'s (2010) meta-analysis (d = 0.62) and Carter and colleagues' (2015) revision of the Hagger et al. meta-analysis in which 41% of the included studies were unpublished (g = 0.43), and bias-corrected meta-analytic estimates such as Carter et al.'s trim-and-fill analysis (g = 0.24). However, the overall effect size of the present replications closely mirrors the regression-based estimate using the precision effect estimation with standard error technique reported by Carter et al. (g = 0.003). The results are consistent with a null effect for ego depletion for the current paradigm. There was substantial heterogeneity in the effect size across labs. A moderator analysis with laboratory language (English-speaking labs vs. non-English-speaking labs) revealed small differences in the effect across English-speaking and non-English-speaking labs, with the 95% confidence intervals for the ego-depletion effect in both groups encompassing zero with substantial heterogeneity.

Unedited Manuscript

An important issue in depletion experiments using sequential-task paradigms, including the present study, 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. Indication of the extent of depletion after the first task is typically inferred from measures that assess 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 depletion version of the task was more demanding and evoked greater perceptions of effort, difficulty, and frustration than the no depletion version. This evidence provides some indication that the initial task was more demanding for participants allocated to the depletion condition relative to controls.

Do the current results suggest that the ego-depletion effect does not exist after all? Certainly the current evidence does raise considerable doubts given the close correspondence of the protocol to the standard sequential-task paradigm typically used 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 bias-uncorrected meta-analytic estimates, are inflated. In a recent commentary, Inzlicht, Gervais, and Berkman (2015) suggested that a range of estimates of the ego-depletion effect size derived from different meta-analytic estimation methods should be considered including: (a) the regression-based estimates reported by Carter et al.; (b) the effect sizes derived from the studies with the top ten largest sample sizes in the meta-analyses; and (c) 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

Unedited Manuscript

the very least, the bias-uncorrected effect size estimates derived from meta-analyses are likely to be substantially inflated. Furthermore, given the rigor with which the current replication was conducted, substantial weight should be attributed to its findings in such considerations.

A number of limitations that may affect interpretation of the effect size generated in current analysis. While the tasks adopted in the current replication closely mirror those that have been used in previous ego-depletion experiments, they are not direct adaptations. For example, the depletion version of the letter 'e' task did not include an initial period where individuals familiarize themselves with the no depletion version of the task used in the control group prior to engaging the depletion version. The initial period is supposed to induce a 'habitual' response that participants would need to override when engaging in the more demanding depletion version (e.g., Fennis, Janssen, & Vohs, 2009). It could therefore be argued that the depletion version of the letter 'e' task was not sufficient in inducing a response that had to be suppressed by participants, that is, suppressing the urge to respond to a letter 'e' in favor of applying the conditional rules. However, in addition to Sripada et al.'s study, a number of sequential-task paradigm 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 habit-formation 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 to make 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.

It is possible that the letter 'e' task was sufficiently arduous but not of sufficient duration to deplete individuals' self-control resources. This is consistent with some preliminary

Unedited Manuscript

evidence that task duration moderates the ego-depletion effect (Hagger et al., 2010a), although there is also evidence that longer duration may enhance self-control (Dang, Dewitte, Mao, Xiao, & Shi, 2013). In the current replication, the duration of the letter 'e' task was identical to the task used by Sripada et al., who found it sufficient to induce depletion. Furthermore, the initial task duration of less than 10 minutes used in the current replication is typical in sequential-task experiments (Hagger et al., 2010a). Nevertheless, duration on the first task may serve to moderate ego-depletion and is in keeping with the premise that individuals need to engage in a sufficient period of effortful self-control to induce a depleted state. Future research that systematically varies the duration of the initial task may be informative as to whether task duration can account for variation in ego-depletion findings (Lee, Chatzisarantis, & Hagger, 2016).

The MSIT used as the dependent self-control task here, while fit for purpose as a response inhibition task that has been used previously in sequential-task paradigm experiments, also led to a high number of participant exclusions due to low accuracy. Although the instructions focused on the importance of correct responses, participants were also told to "go as quickly as you can", so it may be that some participants may have attached high value to rapid responses over correct answers when responding, resulting in a speed-accuracy trade-off. However, the exclusion rate in the depletion group was not significantly greater than the rate in the no-depletion group, allaying concerns of bias as a result of greater error rates in the depletion group. Another concern is that participants excluded for low accuracy on the MSIT task might have been more vulnerable to depletion. However, our overall results do not differ when the participants with accuracy rates below criterion levels are included (see Appendix B).

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 sequential-task paradigm. This is consistent with the underlying

Unedited Manuscript

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. While the current replication of the effect using a standardized paradigm and two impulse control tasks provides good evidence of a null ego-depletion effect, further coordinated replication efforts adopting different tasks from multiple domains would provide additional converging evidence that the depletion effect is null, a position that has been advocated elsewhere (Carter et al., 2015; Hagger & Chatzisarantis, 2014).

Finally, we note the non-trivial, moderate levels of heterogeneity in the ego-depletion effect size across laboratories that cannot be attributed to sampling error alone. This is indicative of some instability in the effect size across labs. One possible cause of the heterogeneity is the presence of moderators. For example, cultural differences of participants from the different national groups may have influenced responses to the tasks, perhaps, for example, influencing the amount of effort that participants invested in the tasks. It is also possible that the implementation of the experimental procedure varied across the labs, the stringent specification of the experimental protocol and methods notwithstanding. Previous multi-lab registered replication reports also observed substantial heterogeneity in some, but not all, of the meta-analyses of the replicated effects across labs (Eerland et al., 2016; Klein et al., 2014). The presence of substantial heterogeneity in some effects may provide useful information on the replicability of experimental results in psychological science. Pre-registration and strict specification of procedures in replication projects is aimed at restricting method variance across labs. If substantial unattributed variability in effects remain with this level of stringency and control, then without such controls the variability may be more substantive. Journal editors should, therefore, demand the highest levels of clarity of reporting and precision in study descriptions, including making complete materials and data freely

Unedited Manuscript

available, in order to ensure that research findings can be judged appropriately in the context of the methods used and that results can be replicated with the highest possible levels of precision.

Conclusion

Results from the current multi-lab registered replication of the ego-depletion effect provide evidence that, if there is any effect, it 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 et al., 2010a). While the current analysis provides robust evidence that questions the strength of the ego-depletion effect and its replicability, it may be premature to reject the ego-depletion effect altogether based on these data alone. Of course, the current replication provides an important source of data with regards to the effect given it is based on a pre-registered design with data from multiple labs, but we recognize it is only 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 exploring the possibility of converging evidence from multiple replication efforts across different depletion domains.

It is also important to note that the current replication speaks little to the underlying mechanism for the ego-depletion effect. Numerous alternative explanations have been proposed that challenge the 'strength' or 'resource depletion' model (Beedie & Lane, 2011; Evans, Boggero, & Segerstrom, 2015; Giacomantonio, Jordan, Fennis, & Panno, 2014; Inzlicht & Schmeichel, 2012; Inzlicht, Schmeichel, & Macrae, 2014; Kotabe & Hofmann, 2015) and research identifying the underlying neural processes may shed light on the processes that underpin ego-depletion (Heatherton & Wagner, 2011; Hedgcock, Vohs, & Rao, 2012; Inzlicht & Gutsell, 2007; Kool, McGuire, Wang, & Botvinick, 2013; Loftus, Yalcin, Baughman,

Unedited Manuscript

Vanman, & Hagger, 2015; Schel, Ridderinkhof, & Crone, 2014). 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. The literature on mental fatigue, for example, suggests that self-regulatory failure is a real phenomenon, but may take longer to materialize. This may tally with findings of the current replication which revealed a null meta-analytic effect size of depletion condition on subjective measures of fatigue across studies. Although we note that fatigue ratings were uncorrelated with the ego-depletion effect size for RT and RTV across studies, a lack of an effect of depletion on fatigue may indicate that although the task was sufficiently arduous, as indicated by difficulty, effort, and frustration ratings, it may not have been of sufficient duration or intensity to result in sufficient fatigue, a candidate proxy measure of depletion. We call for further co-ordinated research programmes and syntheses that explore the possible mechanisms for the effect and, particularly, moderating variables and parameters of the sequential task paradigm that may explain variability in depletion effect sizes across studies (Lee et al., 2016), and analogs that may assist in mechanistic explanations for the effect.

Unedited Manuscript

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. Finally, we thank the following people for their help in conducting the experiments: T-Jay Anderson, Dayna Bell, and Kayla Douglas (Birt Lab); Koen Grootswagers, Femke Kortekaas, David Lacle, Geert Telkamp, Danielle van Dijl, Joeri Wissink, and Joey Zagers (Brandt Lab); Patrick Alarcon, Jessee Marriott, Derrick Ocampo, Briana Peralta, and Rachael Van Gundy (Calvillo Lab); Jantine van Soolingen (Cheung Lab); Josh Cook, Anna Finley, Yvette Ibarra, and Laney Rowe

(Crowell Lab); Jessica Carvajal, Giuliana Kunzle, Julio Martin, and Orlando Olano (Evans Lab); Jacqueline Conway and Clarence Kwong (Francis Lab); Felix Burgdorf and Veronika Drößler (Lange Lab); Aza Maltai (Lau Lab); Sophia Huynh, Sarah Kirschbaum, Molly Minnen, and Ana Moldoveanu (Lynch Lab); Camille Piollet (Muller Lab); Adam Burston, Katie Knapp, Randi Nehls, Natalie Nikora, and Olivia Sievwright (Philipp Lab); Emily Devaney, Kaitlin Cassidy, Miriam Mckiney, Caitlin Romano, and Theresa Tokar (Ringos Lab); Isabel van Oorschot and Joyce van Brecht (Schlinkert Lab); Suzanne Bauwens, Tatjana Dessers, Sientje Palmans, and Mitte Scheldeman (Stamos Lab); Marlon Fedke, Jessika Fuhr, Lisa Häfker, Richard Heinrich, and Georg Hetland (Wolff Lab).

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 *E-Prime* licenses was provided to individual labs by the Association for Psychological Science via a grant from the Center for Open Science.

Unedited Manuscript

References

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., . . . Zwaan, R. A. (2014). Registered Replication Report: Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*, 556-578. doi: 10.1177/1745691614545653

Baumeister, R. F., & Alquist, J. L. (2009). Is there a downside to good self-control? *Self and Identity, 8*, 115-130. doi: 10.1080/15298860802501474

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

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

Beedie, C. J., & Lane, A. M. (2011). The role of glucose in self-control: Another look at the evidence and an alternative conceptualization. *Personality and Social Psychology Review, 16*, 143-153. doi: 10.1177/1088868311419817

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

Dang, J., Dewitte, S., Mao, L., Xiao, S., & Shi, Y. (2013). Adapting to an initial self-regulatory task cancels the ego depletion effect. *Consciousness and Cognition, 22*, 816-821. doi: http://dx.doi.org/10.1016/j.concog.2013.05.005

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

de Ridder, D. T. D., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumeister, R. F. (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to

Unedited Manuscript

a wide range of behaviors. *Personality and Social Psychology Review, 16*, 76-99. doi: 10.1177/1088868311418749

Dewitte, S., Bruyneel, S. D., & Geyskens, K. (2009). Self-regulating enhances self-regulation in subsequent consumer decisions involving similar response conflicts. *Journal of Consumer Research, 36*, 394-405. doi: 10.1086/598615

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. doi: 10.1177/0146167208330855

Eerland, A., Sherrill, A. M., Magliano, J. P., Zwaan, R. A., Arnal, J. D., Aucoin, P., . . . Prenoveau, J. M. (2016). Registered replication report: Hart & Albarracín (2011). *Perspectives on Psychological Science, 11*, 158-171. doi: 10.1177/1745691615605826

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

Giacomantonio, M., Jordan, J., Fennis, B. M., & Panno, A. (2014). When motivational consequences of ego depletion collide: Conservation dominates over reward-seeking. *Journal of Experimental Social Psychology, 55*, 217-220. doi: 10.1016/j.jesp.2014.07.009

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. (2010a). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*, 495-525. doi: 10.1037/a0019486

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010b). Self-regulation and self-control in exercise: The strength-energy model. *International Review of Sport and Exercise Psychology, 3*, 62-86. doi: 10.1080/17509840903322815

Heatherton, T. F., & Wagner, D. D. (2011). Cognitive neuroscience of self-regulation failure. *Trends in Cognitive Sciences, 15*, 132-139. doi: 10.1016/j.tics.2010.12.005

Hedgcock, W. M., Vohs, K. D., & Rao, A. R. (2012). Reducing self-control depletion effects through enhanced sensitivity to implementation: Evidence from fMRI and behavioral studies. *Journal of Consumer Psychology, 22*, 486-495. doi: 10.1016/j.jcps.2012.05.008

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)* (September 11, 2015). Available at SSRN: http://ssrn.com/abstract=2659409

Inzlicht, M., & Gutsell, J. N. (2007). Running on empty - Neural signals for self-control failure. *Psychological Science, 18*, 933-937. doi: 10.1111/j.1467-9280.2007.02004.x

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

Unedited Manuscript

Klein, R. A., Ratliff, K. A., Vianello, M., Adams Jr, R. B., Bahník, Š., Bernstein, M. J., . . . Nosek, B. A. (2014). Investigating variation in replicability: A "many labs" replication project. *Social Psychology, 45*, 142-152. doi: 10.1027/1864-9335/a000178

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

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

Lee, N., Chatzisarantis, N. L. D., & Hagger, M. S. (2016). Adequacy of the sequential-task paradigm in evoking ego-depletion and how to improve detection of ego-depleting phenomena. *Frontiers in Psychology, 7*, 136. doi: 10.3389/fpsyg.2016.00136

Loftus, A. M., Yalcin, O., Baughman, F. D., Vanman, E. J., & Hagger, M. S. (2015). The impact of transcranial direct current stimulation on inhibitory control in young adults. *Brain and Behavior, 5*, e00332. doi: 10.1002/brb3.332

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

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.

Schel, M. A., Ridderinkhof, K. R., & Crone, E. A. (2014). Choosing not to act: Neural bases of the development of intentional inhibition. *Developmental Cognitive Neuroscience, 10*, 93-103. doi: http://dx.doi.org/10.1016/j.dcn.2014.08.006

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.

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

Sterne, J. A. C., Egger, M., & Davey Smith, G. (2001). Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal, 323*. doi: 10.1136/bmj.323.7304.101

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. doi: 10.1111/j.0022-3506.2004.00263.x

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.

Unedited Manuscript

Wills, T. A., Isasi, C. R., Mendoza, D., & Ainette, M. G. (2007). Self-control constructs related to measures of dietary intake and physical activity in adolescents. *Journal of Adolescent Health, 41*, 551-558. doi: 10.1016/jjadohealth.2007.06.013

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

Unedited Manuscript

Table 1

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

Lab	Country and native language 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)
Sripada et al. (2014) (basis for replication)	USA (English)	26	0	0	0	3	23	0.32 (0.07)	0.96 (0.16)	28	0	0	0	4	24	0.27 (0.05)	0.91 (0.13)
Birt & Muise	Canada (English)	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)
Calvillo & Mills	USA (English)	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)
Carruth & Miyake	USA (English)	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)
Crowell, Finley, & Schmeichel	USA (English)	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 (English)	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)
Francis & Inzlicht	Canada (English)	71	4	13	0	33	23	0.30 (0.08)	0.86 (0.13)	69	0	12	2	28	27	0.32 (0.09)	0.91 (0.15)
Hagger, Chatzisarantis & Zwienenberg	Australia (English)	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)
Lau & Brewer	USA (English)	67	0	0	1	19	47	0.32 (0.08)	0.97 (0.14)	65	0	0	0	13	52	0.31 (0.08)	0.93 (0.13)
Lynch, vanDellen & Campbell	USA (English)	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)
Philipp & Cannon	New Zealand (English)	43	0	0	0	5	38	0.31 (0.08)	0.95 (0.16)	43	0	0	0	6	37	0.31 (0.07)	0.97 (0.13)
Ringos & Carlucci	USA (English)	50	0	0	1	17	32	0.33 (0.06)	0.93 (0.05)	50	0	0	1	13	36	0.30 (0.06)	0.99 (0.01)
Brandt	Netherlands (Dutch)	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)
Cheung, Kroese, Fennis, & de Ridder	Netherlands (Dutch)	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)
Elson	Germany (German))	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)
Lange, Heise & Hoemann	Germany (German)	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)
Muller, Zerhouni & Batailler	France (French)	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)
Otgaar, Martijn, Alberts, Merckelbach, Michirev & Howe	Netherlands (Dutch)	50	0	0	2	23	25	0.28 (0.07)	0.86 (0.12)	50	0	0	0	6	44	0.31 (0.07)	0.93 (0.14)
Rentzsch, Nalis & Schütz	Germany (German)	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 (Dutch)	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)
Stamos, Bruyneel & Dewitte	Belgium (Dutch)	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)

Unedited Manuscript

Ullrich & Primoceri	Switzerland (German)	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)
Wolff, Muzi & Brand	Germany (German)	55	0	0	1	16	38	0.33 (0.07)	0.97 (0.12)	56	0	0	1	6	49	0.30 (0.06)	0.92 (0.12)
Yusainy, Wimbarti, Nurwanti & Anggono	Indonesia (Indonesian)	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)

Note. Labs are grouped by English-speaking and non-English-speaking labs and in alphabetical order by lead author. 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.

Unedited Manuscript

Table 2

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

Dependent variable	*d*	CI_{95}		SE	*Q*	*p*	I^2
		LL	UL				
RTV							
Full sample	0.04	-0.07	0.15	.06	34.42	.045	36.08
English-speaking labs	0.14	-0.02	0.30	.08	14.38	.156	30.45
Non-English speaking labs	-0.04	-0.18	0.10	.08	16.88	.112	34.82
RT							
Full sample	0.04	-0.07	0.14	.05	33.40	.056	34.13
English-speaking labs	0.08	-0.09	0.24	.08	19.17	.038	47.84
Non-English speaking labs	0.00	-0.14	0.15	.07	13.82	.243	20.38
Letter 'e' accuracy	-1.82	-1.98	-1.67	.08	50.65	.001	56.57
Self-report measures							
Effort	0.78	0.63	0.94	.08	66.16	<.001	66.75
Fatigue	0.09	-0.03	0.20	.06	36.76	.025	40.15
Difficulty	1.91	1.70	2.12	.11	90.27	<.001	75.63
Frustration	0.82	0.67	0.98	.08	66.51	<.001	66.92

Note. In all cases number of studies was 23. 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.

Unedited Manuscript

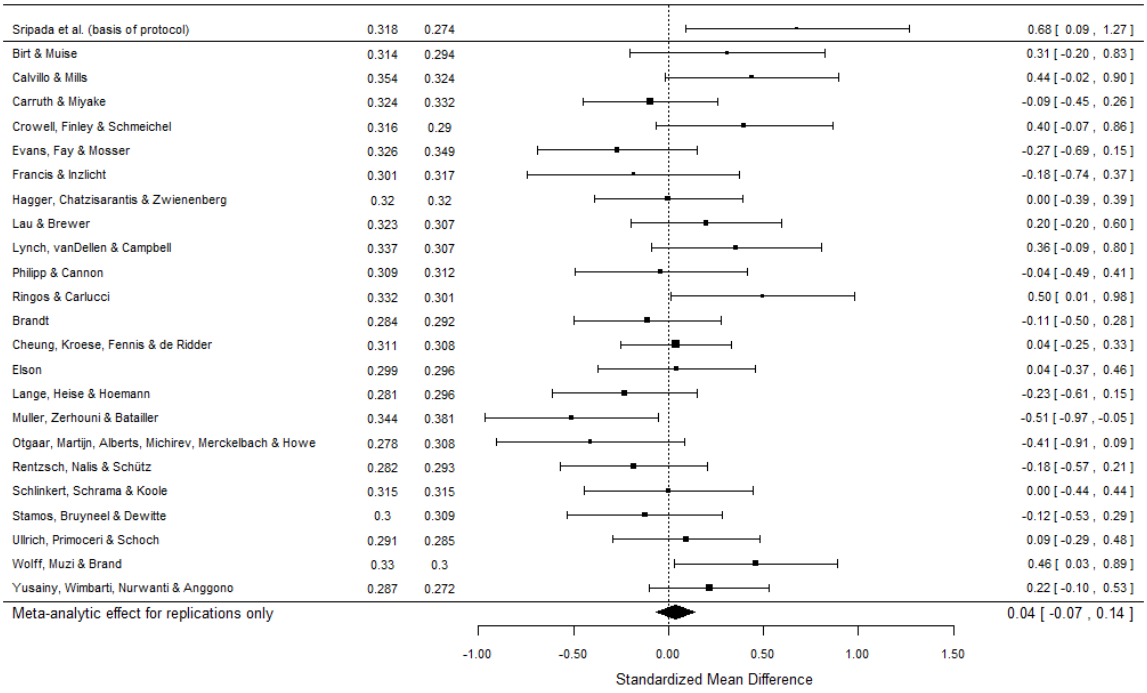

Figure 1. Forest plot of the effect of depletion condition on RTV (reaction time variability) for the multi-source interference task with larger, positive effect sizes indicating greater depletion. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control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

Unedited PDF

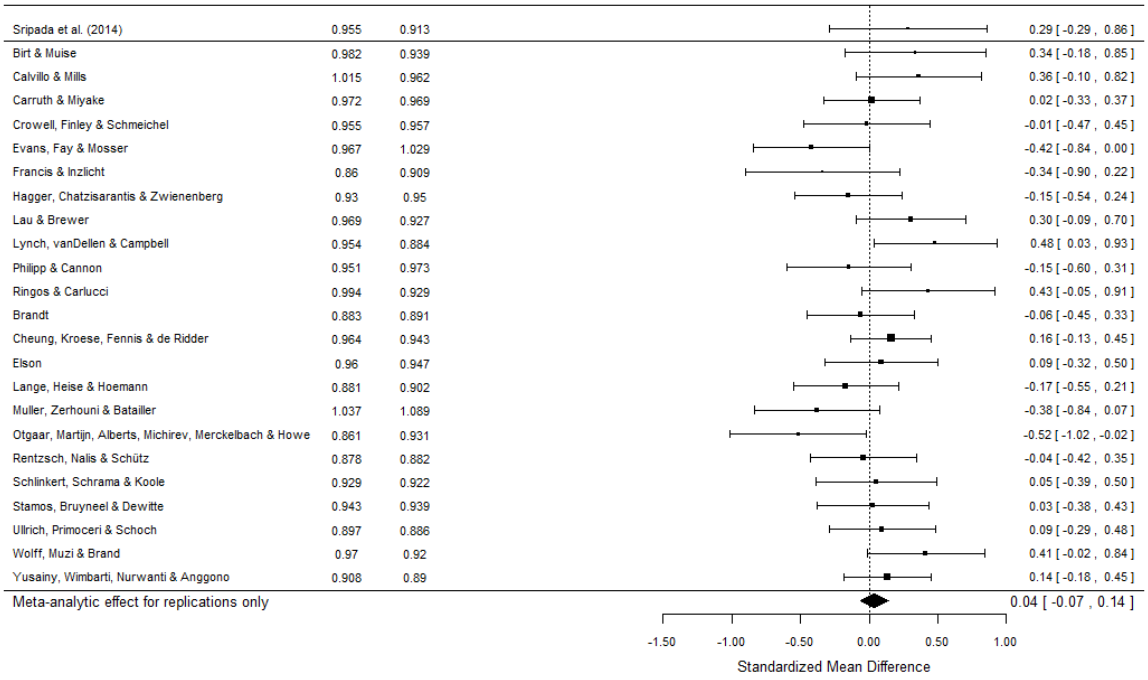

Figure 2. Forest plot of the effect of depletion condition on reaction time (RT) for the multi-source interference task with larger, positive effect sizes indicating greater depletion. For each lab, the figure shows the standardized mean difference (Cohen's *d*) across depletion and control groups and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

Unedited Manuscript