Psychological Research
https://doi.org/10.1007/s00426-019-01230-1

ORIGINAL ARTICLE

Task duration and task order do not matter: no effect on self-control performance

Wanja Wolff[1,2] · Vanda Sieber[3] · Maik Bieleke[4,5] · Chris Englert[2]

Received: 5 February 2019 / Accepted: 6 July 2019

Abstract
The strength model of self-control proposes that all acts of self-control are energized by one global limited resource that becomes temporarily depleted by a primary self-control task, leading to impaired self-control performance in secondary self-control tasks. However, failed replications have cast doubt on the existence of this so-called ego depletion effect. Here, we investigated between-task (i.e., variation in self-control tasks) and within-task variation (i.e., task duration) as possible explanations for the conflicting literature on ego depletion effects. In a high-powered experiment ($N = 709$ participants), we used two established self-control tasks (Stroop task, transcription task) to test how variations in the duration of primary and secondary self-control tasks (2, 4, 8, or 16 min per task) affect the occurrence of an ego depletion effect (i.e., impaired performance in the secondary task). In line with the ego depletion hypothesis, subjects perceived longer lasting secondary tasks as more self-control demanding. Contrary to the ego depletion hypothesis, however, performance did neither suffer from prior self-control exertion, nor as a function of task duration. If anything, performance tended to improve when the primary self-control task lasted longer. These effects did not differ between the two self-control tasks, suggesting that the observed null findings were independent of task type.

Task duration and task order do not matter: no effect on self-control performance

Despite best intentions, self-control does not always work effectively (Vohs & Heatherton, 2000). One of the most popular explanations for this impaired ability to exert self-control has been offered by the strength model of self-control (Baumeister, Vohs, & Tice, 2007). It defines self-control as a volitional act that enables people to regulate certain behavioral tendencies or dominant impulses to accomplish long-term goals (Muraven, Tice, & Baumeister, 1998). For instance, a long-term goal might be to lose weight. Then, self-control is needed to restrain oneself from temptations (e.g., eating a delicious piece of cake) that would lead to immediate joy and gratification but interfere with attaining the long-term weight goal. According to Baumeister et al. (1998), the capacity for such acts of self-control relies on a global, limited resource that is required to regulate all aspects of self-regulatory behavior (e.g., emotion regulation, attention regulation; e.g., Muraven & Baumeister, 2000). Exerting self-control for a certain amount of time is assumed to deplete this resource; and because it is not immediately replenished, performance in subsequent situations that require self-control is impaired. This state of

✉ Wanja Wolff
wanja.wolff@uni-konstanz.de

Vanda Sieber
vanda.sieber@ife.uzh.ch

Maik Bieleke
maik.bieleke@uni-konstanz.de

Chris Englert
christoph.englert@edu.unibe.ch

1 Institute of Sport Sciences, University of Konstanz, Universitätsstraße 10, Konstanz 78464, Germany

2 Department of Educational Psychology, Institute of Educational Science, University of Bern, Fabrikstraße 8, 3012 Bern, Switzerland

3 Institute of Education, University of Zurich, Freiestrasse 36, 8032 Zurich, Switzerland

4 Department of Psychology, University of Konstanz, Universitätsstraße 10, Konstanz 78464, Germany

5 Department of Empirical Educational Research, University of Konstanz, Konstanz, Germany

Published online: 18 July 2019

temporary self-control exhaustion is termed ego depletion (e.g., Baumeister et al., 1998).

To investigate the ego depletion effect, participants first work on a primary task which does (i.e., ego depletion condition) or does not require self-control (i.e., control condition). The subsequent secondary task requires self-control from all participants. It has been repeatedly shown that participants from the depletion condition perform significantly worse in the secondary task compared to participants from the control condition: A substantial body of literature has provided evidence for this ego depletion effect (for a meta-analysis, see Hagger, Wood, Stiff, & Chatzisarantis, 2010). However, failures to replicate the ego depletion effect have accumulated over the years (Lurquin et al., 2016; Xu et al., 2014). In addition, a large registered replication report (RRR) did not find any evidence for the ego depletion effect (Hagger & Chatzisarantis, 2016; for additional analyses of the RRR-data, see Blázquez, Botella, & Suero, 2017; Sripada, Kessler, & Jonides, 2016).

Re-analyses of the most cited meta-analysis (Hagger et al., 2010) on ego depletion suggested that the ego depletion effect might have been overestimated (Carter & McCullough, 2013, 2014). Specifically, these researchers concluded that ego depletion research is affected by publication bias and estimated the true effect-size of ego depletion to be zero (Carter & McCullough, 2014). Support for the notion of a publication bias comes from a recent survey among ego depletion researchers, which revealed that a large portion of ego depletion studies remains unpublished (Wolff, Baumann, & Englert, 2018).

The large-scale replication failure (Hagger et al., 2016) and evidence for a substantial body of gray literature (Wolff et al., 2018) have raised serious doubts regarding the validity of the strength model and caused ongoing discussions about the existence of the ego depletion effect (Hagger et al., 2016; Baumeister & Vohs, 2016). In light of these discussions, it is paramount to investigate possible sources for the inconsistent findings reported in the literature. After all, it is possible that these inconsistencies do not invalidate the strength models propositions, but rather reflect a problematic heterogeneity in the experimental approaches to induce ego depletion (Lee, Chatzisarantis, & Hagger, 2016). Here, we focus on one potential source of the existing inconsistencies that has not yet been systematically investigated: the duration of primary and secondary self-control tasks. Researchers not only use a variety of different self-control tasks (between-task variation; Stroop task, attentional control video; for an overview, see Hagger et al., 2010), they also differ widely in how long participants work on the primary task (within-task variation): For instance, in some studies participants performed more than 200 Stroop trials (Govorun & Payne, 2006), while in other studies participants only had to work on fewer than 50 Stroop trials (Job, Dweck, & Walton, 2010). Task duration is thought to matter, and the duration of the self-control task is assumed to display a linear relationship with the magnitude of the resulting ego depletion effect (Hagger et al., 2010). However, researchers have stressed that it is not clear how long a self-control task must be to induce ego depletion (Lee et al., 2016). Accordingly, a self-control task that is too short might be insufficient for creating detectable levels of ego depletion, leading to the conclusion that no ego depletion effect exists.

The present research

We investigated the role of the duration of primary self-control tasks for ego depletion effects on performance in a subsequent self-control task in a high-powered experiment. Specifically, we assessed the effect of task duration (i.e., 2, 4, 8, and 16 min for each task) on different outcome measures in two ego depletion tasks, namely the Stroop task (Stroop, 1935) and the transcription task (Bertrams, Englert, & Dickhäuser, 2010). The durations were chosen to capture the range of durations most frequently used in ego depletion research (Hagger et al., 2010) and to differentiate this approach from the mental fatigue literature, which mostly uses longer durations (Marcora, Staiano, & Manning, 2009). Both ego depletion tasks are frequently used in ego depletion research (Wolff et al., 2018) and have been reported to be effective at inducing ego depletion (Dang, 2018; Wolff et al., 2018). Moreover, these tasks are particularly well-suited for experimental research: They are easy to standardize to minimize experimenter bias, they can be used as independent (i.e., to deplete self-control strength) and dependent variable (i.e., to measure effects of depleted self-control), and they yield quantitative outcome measures of performance that are easily obtained and interpreted.

For each assessed task duration (i.e., 2, 4, 8, and 16 min), half of the sample worked on the Stroop task first and then on the transcription task, while for the other half of the sample it was the other way around. This non-traditional approach allows for analyzing the effect of each of these two tasks as both an independent variable (i.e., when administered as the primary task) and a dependent variable (i.e., when administered as the secondary task). As both tasks are assumed to require self-control, the strength model predicts that performance on either task should be worse when they are performed as secondary task rather than as primary task (Muraven et al., 1998). In addition, depletion induced by the primary task should be stronger in the experiments with longer task duration, resulting in worse performance in the secondary task. Consequently, if the inconsistent results regarding the ego depletion effect are indeed caused by self-control tasks that were too short, an interaction between-task duration and task order should evince.

Method

Data collection was done online via Amazon Mechanical Turk with the assistance of TurkPrime (Litman, Robinson, & Abberbock, 2017). Studies conducted on Amazon Mechanical Turk, have been shown to give reliable results on different cognitive tasks (Crump, McDonnell, & Gureckis, 2013) including the transcription task and the Stroop task (Savani & Job, 2017). The respondents received monetary compensation for their participation (as the duration of the four experiments differed, the amount of monetary compensation depended on the duration participants had to work on the task: 2 min = 0.50 USD; 4 min = 1.60 USD; 8 min = 2.40 USD; 16 min = 4.0 USD). The study was carried out in accordance with the Helsinki Declaration of 1975 and was approved by the local ethics committee of the University of Bern. The participants who entered the online study were informed about the purpose of the study, delivered informed consent and confirmed that they voluntarily agreed to participate.

Participants

G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) analysis showed that a sample of $N = 675$ was necessary for detecting at least a small to medium effect ($f = 0.16$, $\alpha = 0.05$, $1 - \beta = 0.95$). Out of a total of $N = 975$ participants who started with the task, 729 completed the study. Four subjects had to be excluded because they participated twice and a further 16 had to be excluded because of color blindness. The final sample consisted of $N = 709$ subjects ($n = 333$ female) with a mean age of 36.93 years (SD = 11.03; see Table 1 for detailed descriptive statistics).

Design, procedure, and measures

Participants were randomly assigned to work either on the Stroop task first and then on the transcription task or on the transcription task first and then on the Stroop task. After each self-control task, participants reported their perceived self-control investment and costs. At the end of the experiment, participants provided demographic information (sex, age, color blindness, mother tongue, school degree, ethnic background, and employment status). Finally, participants were probed for suspicion, thanked for their participation, and debriefed.

Measures of perceived self-control investment and costs

In addition to assessing self-control performance, we measured perceived self-control investment and costs. In ego depletion research, this information is usually obtained as a manipulation check to assess if the chosen tasks drew on self-control resources and induced ego depletion. We used single-item measures that have been used in ego depletion research before (Hagger et al., 2016). Specifically, we assessed invested effort (How much effort did you put in the task?) as well as perceived difficulty (How difficult did you find the task?), tiredness (How tired did you feel after doing the task?), and frustration (Did you feel frustrated while you were doing the task?). Each item had to be answered on a 7-point Likert-type scale with specific anchors for effort (1, no effort; 7, a lot of effort), perceived difficulty (1, very easy; 7, very difficult) and identical anchors for tiredness and frustration (1, not at all; 7, very much).

Measures of Self-control performance

The Stroop task (Stroop, 1935) contains a series of color words which are subsequently displayed on the computer screen. The words are either spelled in a color which matches the semantic meaning of the word (e.g., "green" written in green font color; i.e., congruent trial) or in a color which does not match the semantic meaning of the word (e.g., "green" written in blue font color; i.e., incongruent trial). The participants always had to indicate the color in which the word was written, while ignoring the semantic meaning of the respective word by pressing a predefined key on the keyboard. To follow this instruction, participants

Table 1 Descriptive Statistics

Order	Duration	N	Females	Males	Age
Stroop-transcription	2 min	87	$n = 35$	$n = 52$	$M = 37.72$ (SD = 11.18)
Transcription-stroop	2 min	84	$n = 37$	$n = 47$	$M = 39.36$ (SD = 11.14)
Stroop-transcription	4 min	88	$n = 39$	$n = 49$	$M = 36.81$ (SD = 10.62)
Transcription-stroop	4 min	89	$n = 42$	$n = 47$	$M = 35.05$ (SD = 9.53)
Stroop-transcription	8 min	93	$n = 35$	$n = 58$	$M = 35.76$ (SD = 10.22)
Transcription-stroop	8 min	89	$n = 41$	$n = 48$	$M = 38.53$ (SD = 12.51)
Stroop-transcription	16 min	89	$n = 55$	$n = 33$[a]	$M = 36.90$ (SD = 11.10)
Transcription-stroop	16 min	90	$n = 49$	$n = 41$	$M = 35.57$ (SD = 10.44)

[a]One participant chose the "other" option in the Gender question

have to volitionally suppress their dominant word-reading tendency and have to identify the font color instead. The instruction was to correctly identify as many Stroop words as fast as possible. The order of the Stroop trials was randomized and contained the same amount of congruent and incongruent trials. The number of correctly classified congruent and incongruent Stroop trials, as well as the response latencies for the congruent and the incongruent Stroop trials, were measured. Error rates and response latencies were analyzed as dependent variables. As an additional dependent variable, we calculated the Stroop index of interference by subtracting the mean response latency for congruent trials from the mean response latency for incongruent trials (for this procedure, see Richeson & Shelton, 2003). Higher scores on this index indicate higher degrees of interference of the semantic meaning on the color-naming response, meaning worse performance.

In the transcription task (Bertrams et al., 2010), participants had to transcribe a neutral text using the keyboard. The text was displayed on the left side of the screen, while the text field for transcribing the text was displayed on the right side of the screen. The questionnaire was programmed in a way that made copying unavailable. The participants were instructed to never use the letter "e"/"E" and "space bar" while typing. Given that "e"/"E" is the most common letter in the English language, individuals had to volitionally change their dominant writing habits (e.g., Wolff, Baumgarten, & Brand, 2013). The total number of transcribed characters served as the dependent variable.

Statistical approach

All data analyses were conducted with R (3.5.0; R Core Team, 2018). Data organization and visualizations were done with functionality of the TIDYVERSE package (1.2.1; Wickham, 2017) and the COWPLOT package (0.9.4; Wilke, 2019). As manipulation checks, we assessed the effect of performing the self-control tasks on perceived self-control investment (effort) and costs (difficulty, tiredness, and frustration) with 4 (duration: 2 min vs. 4 min vs. 8 min vs. 16 min) × 2 (order: first task vs. second task) analyses of variance (ANOVAs). Separate ANOVAs were run on questions pertaining to the Stroop task and transcription task. Regarding performance, we followed common standards in ego depletion research and analyzed performance in the self-control tasks in a block-wise fashion: To assess Stroop performance (i.e., Stroop interference, mean reaction time in congruent block, mean reaction time in incongruent block, total error rate, error rate in congruent trials, and error rate in incongruent trials) and transcription task performance (overall word count, words transcribed per minute), we conducted 4 (duration: 2 min vs. 4 min vs. 8 min vs. 16 min) × 2 (Order: Stroop-transcription vs. transcription-Stroop) Analyses of Variance (ANOVA). Analyses were done with the AFEX (0.20–2; Singman, Bolker, Westfall, & Aust, 2018) package. To assess difference between specific factor levels, we computed Tukey-corrected post hoc tests with the package EMMEANS (1.3.1; Lenth, 2018). Statistical significance was set at $\alpha = 0.05$ and partial η^2 were calculated as effect-size estimates, where $\eta^2 > 0.01$ represents a small, $\eta^2 > 0.06$ a moderate, and $\eta^2 > 0.14$ a large effect (Cohen, 2013).

Results

Perceived self-control investment and costs

ANOVAs on the effort participants reported to have invested into the Stroop task and into the transcription tasks revealed no significant main effects for order or duration and no significant order × duration interaction, $ps > 0.12$ (Fig. 1, Panel a). Thus, the amount of effort participants were investing into the experimental tasks was not affected by prior self-control exertion, nor by the duration the experimental tasks.

ANOVAs on the perceived difficulty of the Stroop task and the transcription task revealed significant main effects for duration (Stroop task: $F(3, 701) = 3.64$, $p = 0.01$, partial $\eta^2 = 0.02$; transcription task: $F(3, 701) = 4.61$, $p < 0.01$, partial $\eta^2 = 0.02$). For order, a significant main effect emerged only for the Stroop task ($F(3, 701) = 28.14$, $p < 0.01$, partial $\eta^2 = 0.04$) but not for the transcription task ($F(3, 701) = 2.83$, $p = 0.09$, partial $\eta^2 = 0.004$). The ANOVAs revealed no significant order × duration interactions, $ps \geq 0.64$. This indicates that both tasks were perceived as being more difficult when they had to be performed after a first self-control task (Fig. 1, Panel b). The effect sizes further indicate that the perceived difficulty of the Stroop task was more affected by a primary transcription task than the perceived difficulty of the transcription task was affected by a primary Stroop task. Post hoc tests on the effect of duration on perceived difficulty showed that the tasks were perceived as more difficult if they lasted longer. This effect evinced earlier for the transcription task, as indicated by significant differences in difficulty ratings for the comparisons 2-min vs. 16-min ($p = 0.01$), 4-min vs. 16-min ($p = 0.01$). With regard to perceived Stroop difficulty, significant comparisons were 4-min vs. 8-min ($p = 0.03$) and 4-min vs. 16-min ($p = 0.02$).

ANOVAs on how tiring the Stroop and the transcription task were perceived revealed significant main effects for duration (Stroop task: $F(3, 701) = 25.68$, $p < 0.01$, partial $\eta^2 = 0.10$; transcription task: $F(3, 701) = 42.01$, $p < 0.01$, partial $\eta^2 = 0.15$) and order (Stroop task: $F(3, 701) = 19.42$, $p < 0.01$, partial $\eta^2 = 0.03$; transcription task: $F(3, 701) = 13.14$, $p < 0.01$, partial $\eta^2 = 0.02$), but no significant order × duration interaction, $ps \geq 0.38$. Thus, both tasks were perceived as more difficult if they had to be

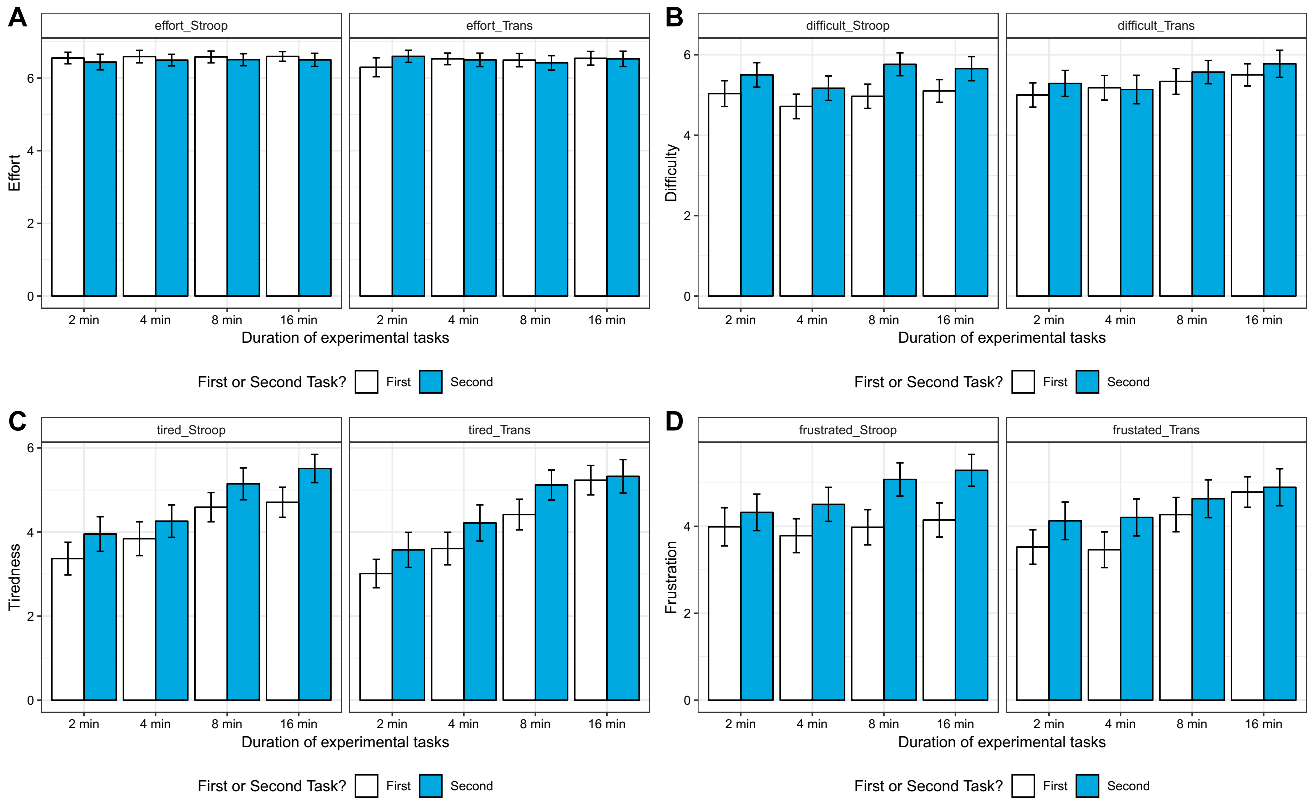

Fig. 1 Perceived self-control demands as a function of task order and duration. Error bars represent 95% confidence intervals. In **a**, participants responded to the question "How much effort did you put in the task?"; in **b** to the question "How difficult did you find the task?"; in **c** to the question "How tired did you feel after doing the task?"; and in **d** to the question "Did you feel frustrated while you were doing the task?"

performed after a first self-control task (Fig. 1, Panel c). Post hoc tests showed that longer durations of the Stroop task and the transcription task were perceived as more tiring (with the exception of the 8-min vs. 16-min and the 2-min vs. 4-min comparisons in the conditions where the Stroop task preceded the transcription task all other ten post hoc comparisons where significant at least at $p<0.04$.).

ANOVAs on how much frustration working on the Stroop task and the transcription task elicited revealed significant main effects for duration (Stroop task: $F(3, 701)=4.00$, $p<0.01$, partial $\eta^2=0.02$; transcription task: $F(3, 701)=11.69$, $p<0.01$, partial $\eta^2=0.05$) and order (Stroop task: $F(3, 701)=33.80$, $p<0.01$, partial $\eta^2=0.05$; transcription task: $F(3, 701)=9.74$, $p<0.01$, partial $\eta^2=0.01$), but no significant order × duration interaction, $ps \geq 0.15$. Thus, both tasks elicited more frustration if they had to be performed after a first self-control task and when they had to be performed longer (Fig. 1, Panel d).

Although the interaction of order × duration on the frustration elicited by the Stroop task failed to reach statistical significance, visual inspection of the interaction suggests that the increase in frustration as a function of task duration appears to occur primarily when the Stroop task was performed after the transcription task. Indeed, post hoc tests revealed no significant differences in frustration as a function of task duration, when the Stroop task was performed as a first task, all $ps>0.58$. However, when the Stroop was performed as the second task, it was perceived as being more and more frustrating as the task got longer. This is underlined by significant differences in the 2-min vs. 8-min ($p=0.04$), the 2-min vs. 16-min ($p<0.01$), and the 4-min vs. 16-min ($p=0.03$) comparisons. No such differentiation was evident for the transcription task (interaction: $p=0.43$). Here, post hoc tests showed that—irrespective of order—longer task duration elicited more frustration. This is underlined by significant differences in the 2-min vs. 8-min ($p=0.01$), the 2-min vs. 16-min ($p<0.01$), the 4-min vs. 8-min ($p=0.01$), and the 4-min vs. 16-min ($p<0.01$) comparisons.

Self-control failures: stroop performance

Response times

ANOVAs on the Stroop interference score revealed a significant main effect for duration ($F(3, 701)=4.75$,

$p < 0.01$, partial $\eta^2 = 0.02$) but neither for order, nor for the order × duration interaction, $ps \geq 0.30$ (*Fig*. 2, panel a). Thus, Stroop interference was not affected by a prior completion of the transcription task. Post hoc tests on the effect of duration revealed that the Stroop interference in the 2-min condition was significantly higher than in the 4-min ($p = 0.03$), 8-min ($p < 0.01$), and 16-min ($p < 0.01$) conditions. No other differences were significant. Thus, longer experimental duration led to an improved performance on the Stroop task. A ceiling of performance improvement was reached already after 4 min and from then on, no further improvements occurred.

For reaction times in the incongruent and congruent blocks, the statistical analyses yielded similar results (Fig. 2, Panels b and c). Main effects of duration (incongruent trials: $F(3, 701) = 10.93$, $p < 0.01$, partial $\eta^2 = 0.04$; congruent trials: $F(3, 701) = 8.31$, $p < 0.01$, partial $\eta^2 = 0.03$) were significant, but neither were the main effects for order or the order × duration interaction, $ps > 0.30$. Post hoc tests again revealed that the effect of order can be ascribed to inferior performance in the 2-min condition compared to the other conditions, $ps < 0.01$. We observed no differences between 4-min, 8-min, and 16-min, respectively, $ps \geq 0.91$.

Errors

ANOVAs on the overall error rate and the error rate in the congruent blocks revealed no significant main effects for duration and order and no order × duration interaction, $ps \geq 0.13$ (Fig. 2, Panels d and f). However, the ANOVA on the error rate in the incongruent block revealed a significant effect of duration $F(3, 701) = 4.02$, $p < 0.01$, partial $\eta^2 = 0.02$), but again no effect of order and no order × duration interaction, $ps \geq 0.33$ (Fig. 2, panel e). Thus, none of the error measures were affected by prior completion of the transcription task.

Post hoc tests revealed a significantly reduced error rate in the 4-min condition compared to the 2-min condition, $p < 0.01$. Although error rates in the 8-min ($p = 0.22$) and the 16-min ($p = 0.14$) conditions were descriptively lower than the 2-min condition, these differences did not reach significance. All other comparisons were not significant, $ps \geq 0.41$. Thus, only the error measure in the incongruent block, i.e., when the task is most difficult, was affected by the duration of the task. In line with the results for the reaction time-based performance measures, performance appears to improve and reach a ceiling quite rapidly.

Self-control failures: transcription task performance

ANOVAs on the number of words transcribed revealed a significant effect for duration ($F(3, 701) = 308.28$, $p < 0.01$, partial $\eta^2 = 0.57$), but not for order or the order × duration interaction, $ps \geq 0.12$ (Fig. 3, Panel A). Expectedly, longer duration of the condition allowed for more words to be transcribed. Again, the number of words transcribed was not affected for subjects who had performed the Stroop task before. To assess if the increase in words transcribed was scaled according to the experimental duration, we ran an ANOVA on the words transcribed per minute. This analysis still revealed a significant main effect for duration ($F(3, 701) = 3.61$, $p = 0.01$, partial $\eta^2 = 0.02$) but not for order or the order × duration interaction, $ps \geq 0.47$ (Fig. 3, panel b). Post hoc tests on the effect of duration on words transcribed per minute showed that participants in the 4-min condition outperformed participants in the 2-min ($p = 0.05$), 8-min ($p = 0.04$), and the 16-min ($p = 0.02$) variants. None of the other comparisons was significant, $ps \geq 0.99$.

Discussion

We investigated the effect of performing a primary self-control task on performance in a subsequently performed secondary self-control task. Participants were randomly assigned to an order in which the two self-control tasks were to be performed. The duration of primary and secondary tasks was varied (2, 4, 8, or 16 min per task), to assess the effect of prolonging self-control exertion on performance in a secondary self-control task. Contrary to the proposition of the strength model of self-control (Muraven et al., 1998), performance did neither suffer in response to prior self-control exertion, nor as a function of task duration. If anything, results even point to the contrary: performance tended to improve when the primary self-control task was of longer duration. Further, we did not observe any significant duration × order interactions, which suggests that failures to find impaired performance after prior self-control exertion is not the result of too short primary tasks. In addition, effects did not differ between the two self-control tasks (i.e., Stroop task and transcription task), which suggests that the observed null findings did also not hinge on one badly chosen type of task.

In line with the behavioral data, our results regarding the manipulation checks—perceived self-control investment and costs—suggest that participants invested similar effort in the two tasks irrespective of how long they were or if they had already performed the respective other self-control task. This investment came, however, with perceived costs and these costs were scaled along task duration and prior self-control exertion. Thus, participants experienced the tasks as more difficult, tiring and frustrating when they had to be performed longer or after a primary self-control task. These effects were consistent across self-control tasks.

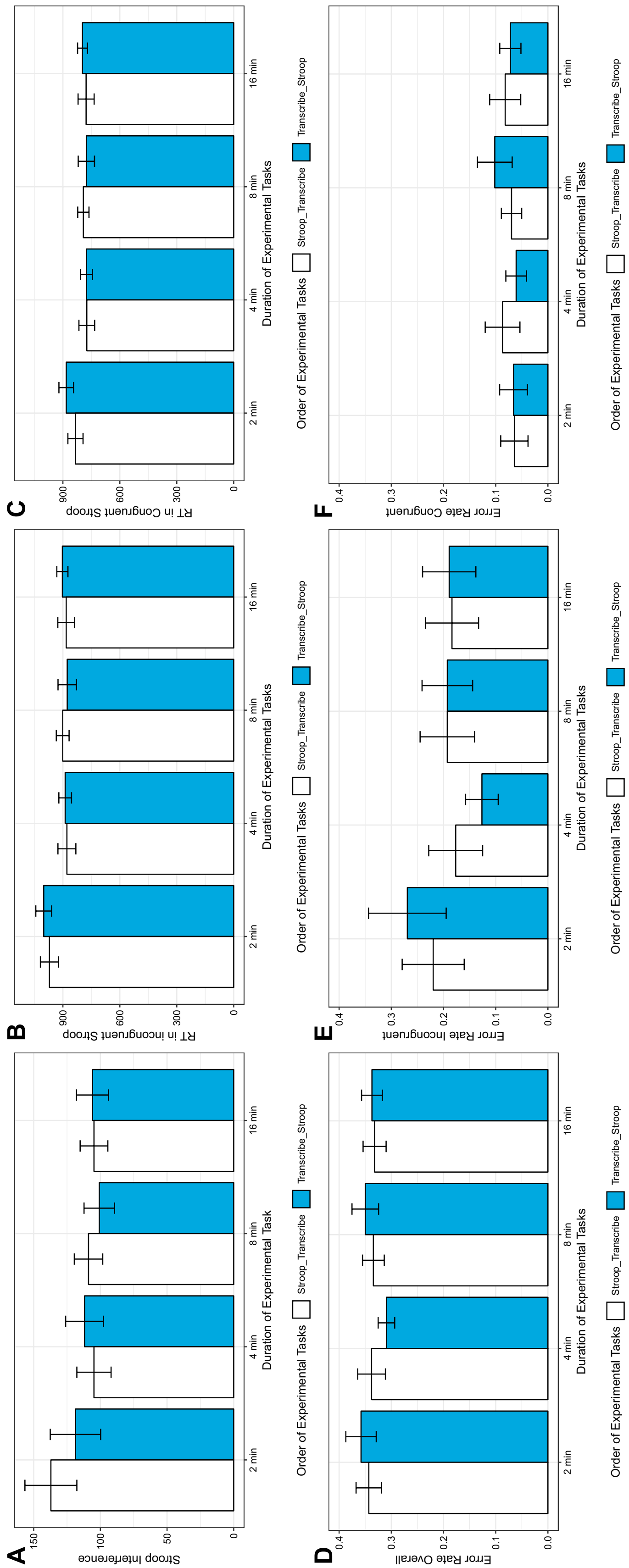

Fig. 2 Stroop performance as a function of task duration and task order. Error bars represent 95% confidence intervals

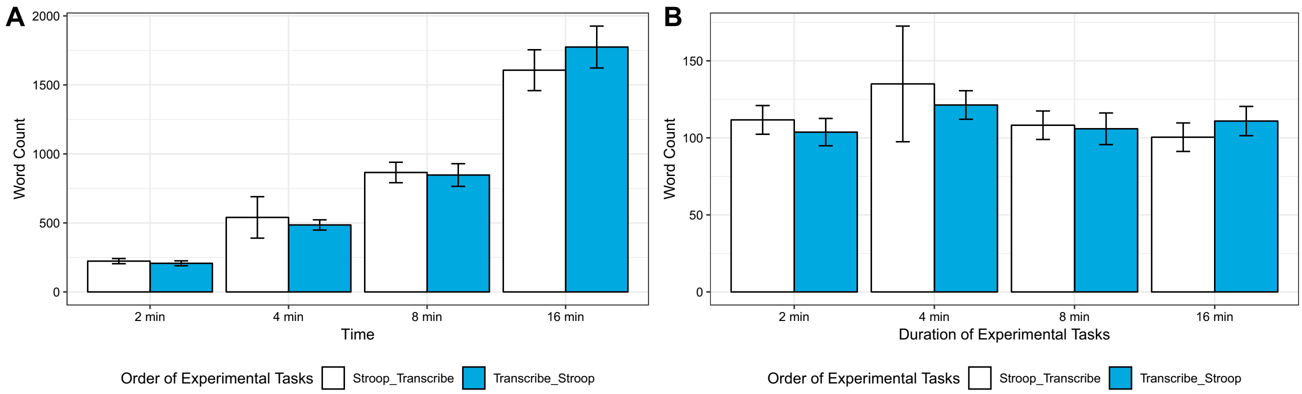

Fig. 3 Performance in the transcription task as a function of task duration and task order. Error bars represent 95% confidence intervals

Implications

In the present research, prior self-control exertion and prolonged task duration did not affect performance on two widely used self-control tasks. However, prolonged task duration and prior self-control exertion resulted in a rise of perceived self-control costs, while the perceived investment of effort stayed on the same level. Thus, in terms of performance, our results do not support the strength model of self-control (Baumeister et al., 2007; Muraven et al., 1998). In terms of subjective experience, however, they are in line with the models' propositions. These results have important implications for the concept of ego depletion and for research on self-control in general. Below we address three tentative interpretations of our findings: self-control is not a limited resource, learning and boredom might modulate the self-control demands induced by a task, and objective performance is no valid indicator for self-control costs.

Does self-control rely on a limited resource?

Our findings regarding overt performance are difficult to reconcile with the predictions of the strength model. They are more in line with recent large-scale replication failures (Hagger et al., 2016) and evidence for publication bias in the literature on ego depletion (Carter & McCullough, 2014; Wolff et al., 2018). The model proposes a reliance on limited resources, meaning that a depletion of resources should result in decreased performance (Baumeister et al., 2007). The failure to observe this decrease aligns with research challenging the empirical (Vadillo, Gold, & Osman, 2016) and conceptual basis (Beedie & Lane, 2012) of a limited physiological substrate for self-control.

In addition to the idea of resource depletion, alternative theoretical accounts on why the allocation of control is perceived as costly (Kool & Botvinick, 2013; Kurzban, Duckworth, Kable, & Myers, 2013) and why people try to avoid it (Inzlicht, Shenhav, & Olivola, 2018) have been proposed (for an overview, see Shenhav et al., 2017). One explanation is that control is perceived as costly to avoid cross talk, which occurs when multiple processes compete for the same neural representations and thereby create a local bottleneck for information processing (Shenhav et al., 2017). Systems that rely on shared neural representations allow for fast and efficient learning and abstract inference (Musslick et al., 2016; Shenhav et al., 2017). However, the shared use of representations severely limits a systems capacity for controlled processing (Feng, Schwemmer, Gershman, & Cohen, 2014). According to this line of thought, exertion of control is perceived as costly not because a resource is depleted, but because exertion of control might prevent the exertion of a concurrent control command (Shenhav et al., 2017). Thus, the perceived costs of control signal the opportunity costs of continuing a chosen course of action (Kurzban et al., 2013). From this perspective, our results can be readily explained: Prior self-control exertion and increased task duration led to increased perceptions of costs, while the self-reported effort stayed the same. Consequently, no decrease in performance was observed.

Task-induced self-control demands might change over time

Another interpretation of our findings might be that the self-control demands that are imposed by a task might change when the duration of the task is varied: In the Stroop task, participants tended to commit fewer errors and to respond faster when the task lasted longer. Thus, an increase in speed was not traded off against accuracy. This highlights an important point, which we believe has not received sufficient attention in the ego depletion literature: An initially difficult and self-control demanding task might lose these characteristics due to learning. Already in his now classic experiment, Stroop showed how an initially control demanding color-naming task could be performed faster after learning

(MacLeod, 1991; Stroop, 1935). Importantly—and in line with the idea of cross talk prevention—, learning leads to a greater automatization of behavior, which is accompanied by a separation of initially shared neural representations (Garner & Dux, 2015). Such distinct representations allow for parallel processing, thereby reducing the self-control demands compared to when a task is executed using shared representations.

To complicate matters further, a task that was initially challenging might become boring after prolonged execution. Although an easier task supposedly incurs less costs for control, boredom is thought to signal low reward for a current course of action (Inzlicht et al., 2018). Boredom is a dynamic state (Mills & Christoff, 2018) that impacts sustained attention and is linked with committing more errors when sustained attention is required (Eastwood, Frischen, Fenske, & Smilek, 2012). The effect boredom has on attention is important because it has been proposed that "Attention control is the single most important or influential form of self-control (…)" (Schmeichel & Baumeister, 2010, p. 31). When dynamic effects of learning and boredom on self-control demands cannot be tracked, it is difficult to predict how an increase in task duration affects resource depletion.

Objective performance might not be a valid indicator of self-control costs

In the present study, subjective ratings of self-control costs and objective performance followed different patterns. This is important, because such subjective ratings usually serve as a manipulation check in ego depletion research on whether or not the ego depletion manipulation had actually worked (Hagger et al., 2016). Ours is by no means the first ego depletion study to observe such a disconnect between self-reported perception and observed performance (e.g., Francis, Milyavskaya, Lin, & Inzlicht, 2018) and our findings also align with a large body of literature on cognitive fatigue in neurological patients, where performance and self-report measures of cognitive fatigue repeatedly fail to correlate (DeLuca, 2005; Sandry, Genova, Dobryakova, DeLuca, & Wylie, 2014). Consequently, researchers in this field have questioned the validity of using performance-based measures as indicators for cognitive fatigue because they might fail to validly capture fatigued resources (Sandry et al., 2014). Preliminary evidence for this notion comes from a recent publication, showing that trait self-control predicts the rate of change in cortical and perceptual markers of self-control costs during a fatiguing task, even when controlling for objective performance (Wolff, Schüler, Hofstetter, Baumann, Wolf, & Dettmers, 2019).

Although cognitive fatigue and self-control should not be conflated, similarities on the conceptual and neuronal level have been highlighted recently (Pattyn, Van Cutsem, Dessy, & Mairesse, 2018). It is, therefore, possible that the heterogenous findings on the ego depletion effect might at least partly stem from the inability to assess if the chosen measures capture depletion of resources on a phenomenological level.

Another reason for the disconnect between self-report ratings and objective performance might be that self-control costs were not accurately captured with the chosen single-item measures. Indeed, while these (and similar) items have been repeatedly used as manipulation checks in ego depletion research (e.g., Hagger et al., 2016), validated short-scales for the assessment of self-control costs are currently lacking. However, it has frequently been shown that single-item measures and multiple-item measures perform equally well, and it has been recommended to use single-item measures when the target concept is very concrete (as in our case; Bergkvist & Rossiter, 2007). This is mirrored by the approach chosen in exercise science and in fatigue research, where perceptions of effort are traditionally assessed with single-item measures (e.g., Bieleke & Wolff, 2017; Sandry et al., 2014). Still, we believe it would be helpful to develop and validate a battery of measures that can be used as standards for the assessment of self-control costs.

Conclusion

The present findings are not in line with the assumption of a limited self-control resource which empowers all aspects of self-control. Matters seem to be more complicated, which is why future research should continue to dig deeper into the antecedents of self-control breakdowns. One promising approach might be to focus on potential mediators of the ego depletion effect. For instance, it has been proposed that motivational, emotional and attentional shifts following a first self-control task might lead to self-control impairments in subsequent tasks instead of a depleted resource (Inzlicht & Schmeichel, 2012). As we did not measure motivation, emotion or attention directly in the present study, we would encourage other researchers to replicate our study while also assessing these potential mediators.

Data availability statement The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest with respect to the authorship or the publication of this article. The authors have full control of all primary data and agree to allow the journal to review their data if requested.

Ethical standards This study has been approved by the local ethics committee. All procedures performed were in accordance with the ethical standards of the institutional and/or national research committee and with the 1975 Helsinki declaration and its later amendments. All persons gave their informed consent prior to their inclusion in the study.

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*(5), 1252–1265. https://doi.org/10.1037/0022-3514.74.5.1252.

Baumeister, R. F., & Vohs, K. D. (2016). Misguided effort with elusive implications. *Perspectives on Psychological Science, 11*(4), 574–575. https://doi.org/10.1177/1745691616652878.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16*(6), 351–355. https://doi.org/10.1111/j.1467-8721.2007.00534.x.

Beedie, C. J., & Lane, A. M. (2012). The role of glucose in self-control: Another look at the evidence and an alternative conceptualization. *Personality and Social Psychology Review, 16*(2), 143–153. https://doi.org/10.1177/1088868311419817.

Bergkvist, L., & Rossiter, J. R. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*(2), 175–184. https://doi.org/10.1509/jmkr.44.2.175.

Bertrams, A., Englert, C., & Dickhäuser, O. (2010). Self-control strength in the relation between trait test anxiety and state anxiety. *Journal of Research in Personality, 44*(6), 738–741. https://doi.org/10.1016/j.jrp.2010.09.005.

Bieleke, M., & Wolff, W. (2017). That escalated quickly—planning to ignore RPE can backfire. *Frontiers in Physiology, 8,* 736. https://doi.org/10.3389/fphys.2017.00736.

Blázquez, D., Botella, J., & Suero, M. (2017). The debate on the ego-depletion effect: Evidence from meta-analysis with the p-uniform method. *Frontiers in Psychology, 8,* 2015–2018. https://doi.org/10.3389/fpsyg.2017.00197.

Carter, E. C., & McCullough, M. E. (2013). Is ego depletion too incredible? Evidence for the overestimation of the depletion effect. *Behavioral and Brain Sciences, 36*(6), 683–684. https://doi.org/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. https://doi.org/10.3389/fpsyg.2014.00823.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Abingdon: Routledge.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PLoS One, 8,* 3. https://doi.org/10.1371/journal.pone.0057410.

Dang, J. (2018). An updated meta-analysis of the ego depletion effect. *Psychological Research, 82*(4), 645–651. https://doi.org/10.1007/s00426-017-0862-x.

DeLuca, J. (2005). *Fatigue as a window to the brain*. Cambridge: MIT Press.

Eastwood, J. D., Frischen, A., Fenske, M. J., & Smilek, D. (2012). The unengaged mind: Defining boredom in terms of attention. *Perspectives on Psychological Science, 7*(5), 482–495. https://doi.org/10.1177/1745691612456044.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Feng, S. F., Schwemmer, M., Gershman, S. J., & Cohen, J. D. (2014). Multitasking versus multiplexing: Toward a normative account of limitations in the simultaneous execution of control-demanding behaviors. *Cognitive, Affective and Behavioral Neuroscience, 14*(1), 129–146. https://doi.org/10.3758/s13415-013-0236-9.

Francis, Z., Milyavskaya, M., Lin, H., & Inzlicht, M. (2018). Development of a within-subject, repeated-measures ego-depletion paradigm. *Social Psychology, 49*(5), 271–286. https://doi.org/10.1027/1864-9335/a000348.

Garner, K. G., & Dux, P. E. (2015). Training conquers multitasking costs by dividing task representations in the frontoparietal-subcortical system. *Proceedings of the National Academy of Sciences, 112*(46), 14372–14377. https://doi.org/10.1073/pnas.1511423112.

Govorun, O., & Payne, B. K. (2006). Ego—depletion and prejudice: Separating automatic and controlled components. *Social Cognition, 24*(2), 111–136. https://doi.org/10.1521/soco.2006.24.2.111.

Hagger, M. S., & Chatzisarantis, N. L. D. (2016). Commentary: Misguided effort with elusive implications, and sifting signal from noise with replication science. *Frontiers in Psychology, 7*(October), 255–264. https://doi.org/10.3389/fpsyg.2016.00621.

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A. R., … et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*(4), 546–573. https://doi.org/10.1177/1745691616652873.

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*(4), 495–525. https://doi.org/10.1037/a0019486.

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science, 7*(5), 450–463. https://doi.org/10.1177/1745691612454134.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: effort is both costly and valued. *Trends in Cognitive Sciences, 22*(4), 337–349. https://doi.org/10.1016/j.tics.2018.01.007.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego depletion-is it all in your head? Implicit theories about willpower affect self-regulation. *Psychological Science, 21*(11), 1686–1693. https://doi.org/10.1177/0956797610384745.

Kool, W., & Botvinick, M. (2013). The intrinsic cost of cognitive control. *Behavioral and Brain Sciences, 36*(6), 697–698. https://doi.org/10.1017/S0140525X1300109X.

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences, 36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196.

Lee, N., Chatzisarantis, N., & 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. https://doi.org/10.3389/fpsyg.2016.00136.

Lenth, R. (2018). emmeans: Estimated marginal means, aka least-squares means. R package version 1.3.1. https://cran.r-project.org/web/packages/emmeans/index.html.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433–442. https://doi.org/10.3758/s13428-016-0727-z.

Lurquin, J. H., Michaelson, L. E., Barker, J. E., Gustavson, D. E., Von Bastian, C. C., Carruth, N. P., & Miyake, A. (2016). No evidence of the ego-depletion effect across task characteristics and individual differences: A pre-registered study. *PLoS One, 11*(2), 1–20. https://doi.org/10.1371/journal.pone.0147770.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*(2), 163–203. https://doi.org/10.1037//0033-2909.109.2.163.

Marcora, S. M., Staiano, W., & Manning, V. (2009). Mental fatigue impairs physical performance in humans. *Journal of Applied Physiology, 106*(3), 857–865. https://doi.org/10.1152/japplphysiol.91324.2008.

Mills, C., & Christoff, K. (2018). Finding consistency in boredom by appreciating its instability. *Trends in Cognitive Sciences, 22*(9), 744–747. https://doi.org/10.1016/j.tics.2018.07.001.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin, 126*(2), 247–259. https://doi.org/10.1037/0033-2909.126.2.247.

Muraven, M., Tice, M. D., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*(3), 774–789.

Musslick, S., Dey, B., Özcimder, K., Patwary, M. A., Willke, T. L., & Cohen, J. D. (2016). Parallel processing capability versus efficiency of representation in neural networks. *15th Neural Computation and Psychology Workshop, 1,* 13–14. https://doi.org/10.13140/RG.2.2.30076.54407.

Pattyn, N., Van Cutsem, J., Dessy, E., & Mairesse, O. (2018). Bridging exercise science, cognitive psychology, and medical practice: Is "cognitive fatigue" a remake of "the emperor's new clothes"? *Frontiers in Psychology, 9*((SEP)), 1–13. https://doi.org/10.3389/fpsyg.2018.01246.

R Core Team (2018). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. https://www.r-project.org/.

Richeson, J. A., & Shelton, J. N. (2003). Effects of interracial contact on executive function. *Psychological Science, 14*(3), 1–5. https://doi.org/10.1111/1467-9280.03437.

Sandry, J., Genova, H. M., Dobryakova, E., DeLuca, J., & Wylie, G. (2014). Subjective cognitive fatigue in multiple sclerosis depends on task length. *Frontiers in Neurology, 5,* 1–7. https://doi.org/10.3389/fneur.2014.00214.

Savani, K., & Job, V. (2017). Reverse ego-depletion: Acts of self-control can improve subsequent performance in Indian cultural contexts. *Journal of Personality and Social Psychology, 113*(4), 589–607. https://doi.org/10.1037/pspi0000099.

Schmeichel, B. J., & Baumeister, R. F. (2010). Effortful attention control. In B. Bruya (Ed.), *Effortless attention: A new perspective in the cognitive science of attention and action* (pp. 29–50). Cambridge: MIT Press.

Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience, 40*(1), 99–124. https://doi.org/10.1146/annurev-neuro-072116-031526.

Singman, H., Bolker, B., Westfall, J. & Aust, F. (2018). afex: Analysis of Factorial Experiments. R package version 0.20-2. https://cran.r-project.org/web/packages/afex/index.html.

Sripada, C., Kessler, D., & Jonides, J. (2016). Sifting signal from noise with replication science. *Perspectives on Psychological Science, 11*(4), 576–578. https://doi.org/10.1177/1745691616652875.

Stroop, J. R. (1935). Studies of interference. *Journal of Experimental Psychology, 18*(6), 643–662. https://doi.org/10.1037/h0054651.

Vadillo, M. A., Gold, N., & Osman, M. (2016). The bitter truth about sugar and willpower: The limited evidential value of the glucose model of ego depletion. *Psychological Science, 27*(9), 1207–1214. https://doi.org/10.1177/0956797616654911.

Vohs, K. D., & Heatherton, T. F. (2000). Self-regulatory failure: A resource-depletion approach. *Psychological Science, 11*(3), 249–254. https://doi.org/10.1111/1467-9280.00250.

Wickham, H. (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1. https://cran.r-project.org/web/packages/tidyverse/index.html.

Wilke, C. O. (2019). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 0.9.4. https://cran.r-project.org/web/packages/cowplot/index.html.

Wolff, W., Baumann, L., & Englert, C. (2018). Self-reports from behind the scenes: Questionable research practices and rates of replication in ego depletion research. *PLoS One, 13*(6), 1–11. https://doi.org/10.1371/journal.pone.0199554.

Wolff, W., Baumgarten, F., & Brand, R. (2013). Reduced self-control leads to disregard of an unfamiliar behavioral option. *Substance Abuse Treatment, Prevention, and Policy, 8*(41), 1–6. https://doi.org/10.1186/1747-597X-8-41.

Wolff, W., Schüler, J., Hofstetter, J., Baumann, L., Wolf, L., & Dettmers, C. (2019). Trait self-control outperforms trait fatigue in predicting MS patients' cortical and perceptual responses to an exhaustive task. *Neural Plasticity*. https://doi.org/10.1155/2019/8527203.

Xu, X., Demos, K. E., Leahey, T. M., Hart, C. N., Trautvetter, J., Coward, P., … et al. (2014). Failure to replicate depletion of self-control. *PLoS One, 9*(10), 1–5. https://doi.org/10.1371/journal.pone.0109950.