Author accepted manuscript version: to appear in Social Psychology

Is the letter cancellation task a suitable index of ego-depletion? Empirical and conceptual issues

Marina C. Wimmer[a], Lenard Dome[a], Peter J.B. Hancock[b], & Thomas Wennekers[c]

[a]University of Plymouth, School of Psychology, Cognition Institute, Plymouth, UK, PL4 8AA;

[b]University of Stirling, Psychology, Faculty of Natural Sciences, Stirling, UK, FK9 4LA

[c]University of Plymouth, School of Computing, Electronics, and Mathematics, Cognition Institute, Plymouth, UK, PL4 8AA;

Word Count = 6752 (excl. title page, abstract)

Send correspondence to: Dr. Marina C. Wimmer
School of Psychology
University of Plymouth, Plymouth, UK PL4 8AA
email: marina.wimmer@plymouth.ac.uk
phone: +44/(0)1752 585881
facsimile: +44/(0)1752 584808

Author contribution

All authors were involved in all parts of the research; MW designed the study, collected data, analysed the data and took the main responsibility of the write up of the research, LD collected the data, analysed the data, and contributed towards the write up of the research, PH analysed the data, and contributed towards write up of the research, TW designed the study, analysed the data, and contributed towards the write up of the research.

Abstract

The aim was to quantify ego-depletion and measure its effect on inhibitory control. Adults (N = 523) received the letter “e” cancellation ego-depletion task and were subsequently tested on Stroop task performance. Difficulty of the cancellation task was systematically manipulated by modifying the text from semantically meaningful to non-meaningful sentences and words (Experiment 1) and by increasing ego-depletion rule complexity (Experiment 2). Participants’ performance was affected by both text and rule manipulations. There was no relation between ego-depletion task performance and subsequent Stroop performance. Thus, irrespective of the difficulty of the ego depletion task, Stroop performance was unaffected. The widely used cancellation task may not be a suitable inducer of ego-depletion if ego-depletion is considered as a lack of inhibitory control.

Keywords: ego-depletion, inhibitory control, letter cancellation, replication, conceptual questions

Ego-depletion, the finding that self-control is temporarily impaired because of a previous task that has tapped into self-control resources (Baumeister, Bratslavsky, Muraven, & Tice, 1998) is one of the most studied phenomena in Social Psychology (PsychINFO search term “ego-depletion” yields 518 results). Yet there is current doubt on the strength of the effect or whether the effect exists at all (Carter, Kofler, Forster, McCullough, 2015; Etherton, et al., 2018; Hagger et al., 2016; Lurquin et al., 2016). There is also a conceptual critique that self-control is an undefined measure in ego-depletion (Lurquin & Miyake, 2017). To gain insight into the specific mechanisms underlying ego-depletion and the strength of the effect, the current research systematically manipulated task demands for the ego-depletion task itself and quantified performance to examine subsequent effects on inhibitory control.

The ego-depletion effect works as follows: a task that supposedly taps into self-control resources reduces performance in a subsequent task (the outcome task) requiring self-control (Alós-Ferrer, Hügelschäfer, & Li, 2015; Hagger, Wood, Stiff, & Chatzisarantis, 2010). For example, watching a silent video of an interview with irrelevant words appearing on the screen, participants who are instructed to ignore the words (ego-depletion) are more likely to defer a decision to buy a camera (outcome task) than those who are told to watch the video (control) (Vonash, Vohs, Pocheptsova Gosh, & Baumeister, 2017). One explanation is the “processing limit” energy model, suggesting that self-control is a limited resource, so that if energy is consumed in a self-control task, performance will be reduced in the outcome task (Baumeister, 2014; Baumeister, Tice, & Vohs, 2018). This can also extend to physical aspects of fatigue (Evans, Boggero, & Serstrom, 2016). The model does not assume the depletion of specific cognitive processes but the depletion of general psychological (and physical) processing power. Alternatively, the “high-level” account suggests that when a task requires self-control, participants may subsequently have reduced attention to cues that require control and reduced motivation to exert control (Inzlicht & Schmeichel, 2012; Inzlicht, Schmeichel, & Macrae, 2014). In other words, this account specifically suggests that high-level cognitive processes such as attentional and motivational control are depleted. Thus, the underlying cause of ego depletion is assumed to be “shifts” in attentional and motivational goals (high-level account) as opposed to reduced psychological (and potentially physical) processing power (Baumeister, 2014; Evans et al., 2016; Inzlicht et al., 2014).

While there is empirical support for either theory (e.g., Alós-Ferrer et al., 2015; Muraven & Slessareva, 2002), the strength of the effect has recently been put into question. First, a meta-analysis that only included experiments with frequently used ego-depletion and outcome tasks showed the effect is not different from zero. Second, Bayesian analyses that compared performance in three different pairs of previously widely used ego-depletion and outcome tasks (restricted writing, letter cancellation, Stroop task followed by handgrip duration, anagrams, and mental arithmetic) revealed evidence in favour of the null hypothesis. Third, a preregistered replication study involving 23 laboratories that used the letter cancellation task as ego-depletion measure and the multi-source interference task as outcome measure of inhibitory control (Carter, et al., 2015; Etherton, et al., 2018; Hagger et al., 2016) raised further doubts about the strength of the effect. However, interestingly the replication study showed great variations in task performance in accuracy both between ego depleted and control participants and across different laboratories (15%-44% of participants performing < 80% correct) and a great range of effect sizes on response time differences between ego depleted and control participants (95% CI on Cohen’s d = -0.06, 0.36). Moreover, other studies using the letter cancellation task in single laboratory studies found positive ego-depletion effects using different types of outcome tasks (Baumeister et al., 1998; Stripada, Kessler, & Jonides, 2014; Wan & Sternthal, 2008; see also Hagger et al., 2010 for an overview) perhaps prematurely dismissing the existence of the effect *per se*.

Adding further complexity to the ego-depletion phenomenon is that the self-control process is not conceptualized and the term is loosely applied to a variety of tasks without any conceptual check that the task itself taps into self-control (Lurquin & Miyake, 2017). In fact the type of depletion task, such as controlling emotions, thoughts, impulses, attention, choice and volition, cognitive and social processing, has no effect on the strength of the ego-depletion effect (Haggart et al., 2010). Thus, it is unclear what processes underlie ego-depletion *per se*, while performance on the ego-depletion task itself is largely ignored. The standard approach is to administer a questionnaire on fatigue level as a manipulation check after the ego-depletion task has been completed. There is rarely a quantitative measure of ego-depletion nor is it taken into consideration for subsequent task performance, with an exception being Lurquin et al. (2016) who did not find the effect *per se*. In an attempt to gain insight into what aspects of self-control are depleted we systematically manipulated one of the most widely used ego-depletion tasks, the letter cancellation task, and quantified performance to examine its effects on the most established inhibitory control measure, the Stroop task (Stroop, 1935). We created different versions of the letter cancellation task to quantify ego-depletion as an inducer of self-control, that is, how much self-control is lost under different conditions in terms of difficulty level and to measure its effectiveness for ego-depletion research in the first place.

Arber et al. (2017) show, across four variants of the letter cancellation task, that performance decreases with time spent on the task. A fifth test showed that initial working memory score correlated with performance on the letter cancellation task, but not with the rate of performance decrease with time. Conversely, the rate of decrease on letter crossing did relate to subsequent working memory score, which suggests that some resource, if only willingness to attend to the experiment, is being depleted.

Myers et al. (2018) tested the claim that habit formation followed by inhibition of the habit are required for depletion to occur. Their first experiment used a meaningless sequence of letters and found no decrease in performance with time. Reverting to coherent text produced the same pattern as Aber et al. (2017), with the addition of effects due to the frequency of vowel pairs in the corpus. These may have caused some of the depletion effects observed in the earlier studies. However, it is apparent that manipulations to the semantics of the presented text will have an effect and that prompts the design of our first experiment.

Our experiment 1 aimed to change the difficulty of finding the target letters by manipulating the text. Letter detection is better when text has no semantic meaning; at the sentence level, by scrambling the word order, or at the word level, by scrambling the letter order. Meaningful words can be identified as a whole unit and thus single letters are missed (see e.g., Healy, 1994 for an overview). Our prediction is that the more meaningful the text, the harder the task and so the greater the subsequent degree of ego-depletion.

Experiment 2 manipulated the complexity of the rules for the letter cancellation task, to test the effect of holding complex versus simple rules in working memory. Our prediction is that more complex rules will result in greater ego-depletion.

Experiment 1

In this first experiment, the rules remained constant (ego-depleted: “in a text cross off any “e” if it is not adjacent to another vowel or one extra letter away from another vowel; control: “cross off any “e”) while the text pattern was manipulated ranging from semantically meaningful to non-meaningful text.

Method

Participants

A total of 310 adults (262 females), M = 21 years, SD = 5 years, recruited via the xxx University online participation system took part. Participants received course credit for participation.

Design

The experiment implemented a 2(condition: ego-depleted versus control) x 3(text manipulation: standard text vs. words scrambled vs. characters scrambled) design where both variables were manipulated between subjects. Participants were subjected either to the control or ego-depletion group. After that, all participants received a computerized Stroop task (Stroop, 1935). The whole experiment lasted around 30 minutes.

Materials and Procedure

Participants in the *control condition* were instructed to cross off all occurrences of the letter “e”. Participants in the *ego-depletion condition* were told in addition, to only cross off an “e” if it is not adjacent to another vowel or one extra letter away from another vowel (thus, one would not cross off the “e” in “vowel”) (see Baumeister et al., 1998). Participants were given 8 example words to understand the rules and had the rules written in front of them on a piece of paper.

The task was performed on a typewritten sheet of paper from a neuroscience article (page 2 of Cabeza, Rao, Wagner, Mayer, & Schacter, 2001) in one out of three different formats: Text 1 (standard text, N = 102) comprised the standard text typeset in LaTex (e.g., the first complete sentence read “They were not taking medications and did not have medical conditions that could affect cerebral blood flow.”). Text 2 (words scrambled, N = 107) was manipulated so that all words were scrambled, thus there was no semantic meaning in each sentence (The order of words was a random permutation of the words in Text 1) (e.g., the first complete sentence read: “For consisted voxel items between are order presented associates a estimated list new have HRF pooled were.”). Text 3 (characters scrambled, N = 101) was manipulated so that all characters in the original text were scrambled (keeping the relative number of characters and the distribution of word-lengths intact) (e.g., the first complete sentence read: “Maet eawa ioo oiteej tfbrelnelln hua atd kle ooit lorhord cdsaeeesnh dnet crtue taemrt ieieedec ttsrn taei.”). In all 3 cases the page of the original text used was converted from a pdf to an ASCII text file and then manipulated in Octave (Eaton, Bateman, Hauberg, Wehbring, 2017). The resulting three texts were typeset using LaTeX (https://www.latex-project.org/) as similar to the original text as possible. This included proper capitalisation of words in sentences, a two column layout with paragraphs as in the original, and boldface fonts of paragraph headings. We could not use the same font as the original text, but all three modified texts used the same font, close to the original.

About half of the participants were subjected to the ego-depletion condition (N = 184) and the other half to the control group (N = 126). Numbers are not even as many ego-depleted participants did not follow both ego-depletion rules which was established via visual inspection by the experimenter after each experimental session. Thus, more participants were allocated to the ego-depletion condition as the experiment proceeded. However, as visual inspection is not a sufficiently accurate method of error detection, performance on the letter cancellation task was analysed with a computerized detection program explained below.

Participants in both ego-depletion and control conditions were told that they had 8 minutes to cross off the letters. After 8 minutes they were stopped, to keep completion time constant across groups and avoid time of completion confounding ego-depletion effects (Hagger et al., 2010).

After that, all participants received a computerized version of the Stroop task, containing 100 word reading-, 100 colour naming-, and 100 interference trials, where trial order was counterbalanced between participants, as used in Wimmer et al. (2017). The Stroop task consisted of colour words (red, green, blue, yellow). In the word reading trials the colour word appeared in black font and participants clicked with the mouse on the according colour patch that appeared above in a 2x2 square configuration. In the colour naming trials, a colour patch appeared and participants clicked on the according colour word, written above in black font in the same configuration. In interference trials a colour word appeared in colour (e.g., blue written in red font) and participants clicked on the colour the word was written in. Most (76) interference trials were incongruent (text colour and colour word different), mixed with 24 congruent trials to increase the interference effect. A mixed design was used to increase the interference effect. Each trial started with a fixation cross in the centre of the screen where the mouse was positioned and inter trial interval was 1000 ms. Participants received 10 practice trials before each trial type.

In the letter cancellation task all participants marked characters on a single printed sheet of paper. All sheets were machine-scored and subsequently individually checked and corrected by hand on a computer using a custom made graphical user interface. The tentative location of markers was first detected automatically: Sheets were scanned and processed in Octave (Eaton, et al., 2014) as grey-scale images. Images were transformed geometrically to match with a scan of an unmarked sheet (using affine transformations with spline-interpolation), brightness normalised and subtracted from the unmarked original. The absolute differences were spatially smoothed, after which spatial maxima were detected. These maxima correspond with potential locations of letters marked by the participants. However, because the data shows a tremendous variation due to various sources of noise (marker symbols chosen, size of markers and precision of their placement, notes, scribble, or debris on the sheets, inhomogeneous scanning, imperfect affine transformation, and more) the machine-scored locations needed to be corrected by hand in a second step. This resulted in less than 1 expected mistake on average per sheet. It is not possible to reach a 100% correct rate, because the placement of markers can be ambiguous. Locations of markers were finally compared with correct locations extracted from correctly annotated sheets for each rule. This used nearest neighbour matching with a maximum distance criterion and resulted in total numbers of correctly and wrongly placed markers.

Results

Throughout both experiments Bonferroni corrections and post-hoc tests were applied. In Experiment 1 ego-depletion and Stroop task performance was examined in several univariate ANOVAs with experimental group (ego-depletion vs. control) and text as between participants variables. Data of 23 participants (12 in the control and 11 in the ego-depletion condition) were excluded in the analyses of ego-depletion task performance due to experimenter error, missing the letter cancellation sheets. All data are available here: https://doi.org/10.6084/m9.figshare.7575581

Ego-depletion task performance

Mean performance on the letter e task as a function of condition and text is shown in table 1.

Total marks set. First we investigated how many characters were crossed off during 8 minutes and whether this differed between ego-depleted and control participants and between texts. More characters were crossed off in the control (M = 344) than in the ego-depletion condition (M = 73), $F(1, 281) = 2138.22$, $p < .001$, $\eta_p^2 = .88$. There was also an effect of text, $F(2, 281) = 3.95$, $p = .02$, $\eta_p^2 = .03$ where fewer characters were crossed in the standard text (M = 199) than in the words scrambled text (M = 218, p = .017) and both did not differ from the characters scrambled text (M = 210, ps > .33). There was no interaction, $F(2, 281) = 1.38$, $p = .25$, $\eta_p^2 = .01$.

Proportion correct. To test accuracy of performance, the number of correct characters crossed based on the total marks set were analysed. Participants in the control condition crossed more correct characters (M = .96) than ego-depleted participants (M = .78), $F(1, 281) = 81.59$, $p < .001$, $\eta_p^2 = .23$. There was also an effect of text, $F(2, 281) = 4.32$, $p = .014$, $\eta_p^2 = .03$ that occurred because there were more correct characters crossed when the text had scrambled characters (M = .90) than scrambled words (M = .83, p = .014), while both did not differ from the standard text (*M* = .88, *p*s > .16). Again, there was no interaction, $F(2, 281) = .72$, $p = .49$, $\eta_p^2 = .005$.

Proportion missing. Finally, the number of correct characters that were not crossed from the total mark number was investigated. For example, if someone crossed off 344 characters in total but in between these missed 64 correct characters the proportion missing would be 64/344 = .19. There were fewer missing characters in the control (*M* = .18) than in the ego depletion group (*M* = .43), $F(1, 281) = 69.22$, $p < .001$, $\eta_p^2 = .20$. There was no effect of text, $F(2, 281) = 2.77$, $p = .07$, $\eta_p^2 = .02$ and no interaction, $F(2, 281) = .39$, $p = .68$, $\eta_p^2 = .003$.

Table 1. Mean performance on letter cancellation as a function of condition and text. Standard deviation in parenthesis.

Condition	Text	Total marks set	Proportion correct	Proportion missing
Control	standard text	328 (61)	.98 (.03)	.21 (.13)
	words scrambled	352 (65)	.92 (.05)	.17 (.13)
	characters scrambled	352 (62)	.97 (.03)	.15 (.17)
Ego-depletion	Standard text	69 (31)	.78 (.19)	.49 (.30)
	Words scrambled	84 (40)	.74 (.21)	.42 (.31)
	Characters scrambled	69 (33)	.83 (.19)	.37 (.25)

Stroop performance

Accuracy was approaching ceiling in all three Stroop trials (interference: *M* = 93.2; word reading: *M* = 98.8; colour naming: *M* = 98.8), therefore, no further statistical analysis was conducted (Table 2).

Table 2. Mean accuracies (percentages) for the three types of Stroop tasks as a function of condition and text. Standard deviation is in parentheses.

	Control			Ego depletion		
	Standard	Words Scrambled	Characters Scrambled	Standard	Words Scrambled	Characters Scrambled
Interference	94.7 (15.6)	96.1 (11.9)	91.3 (21.2)	92.2 (20.2)	92.8 (18.1)	93.0 (18.5)
Colour naming	99.6 (.77)	99.4 (.76)	98.6 (5.88)	99.5 (.70)	98.2 (6.78)	98.0 (7.52)
Word reading	99.5 (.80)	99.5 (1.27)	98.6 (5.90)	99.4 (.91)	98.4 (6.72)	97.9 (7.49)
Overall	97.9 (5.15)	98.3 (3.99)	96.2 (6.54)	97.0 (6.73)	96.5 (8.71)	96.3 (9.22)

Examining overall Stroop response time of correct trials as a function of condition and text, there was no difference between ego-depleted (*M* = 970 ms) and control participants (*M* = 969 ms), $F(1, 303) = .003$, $p = .96$, $\eta_p^2 = .00$. There was also no difference between texts (standard: *M* = 977 ms, scrambled words: *M* = 961 ms, scrambled letters: *M* = 970 ms), $F(2, 303) = .31$, $p = .74$, $\eta_p^2 = .002$ and no interaction, $F(2, 281) = .88$, $p = .42$, $\eta_p^2 = .006$.

To further assess whether there is an overall difference between ego-depleted and control participants, we calculated the Bayes factor, using a JZS prior (Rouder, Speckman, Sun, Morey, & Iverson, 2009). The odds, *BF*=10.97, are strong evidence for the null hypothesis.

Isolating the inhibitory component of the task, mean response times of both word reading and colour naming trials were subtracted from the mean of interference trials. There was no difference between ego-depleted (*M* = 185 ms) and control participants (*M* = 211 ms), $F(1, 281) = 2.22$, $p = .14$, $\eta_p^2 = .007$. There was also no effect of text, $F(2, 303) = 2.16$, *p*

= .12, η_p^2 = .014, and no interaction, $F(2, 303) = .27$, $p = .77$, η_p^2 = .002 (Figure 1). The same results were obtained when the Stroop effect was calculated differently by subtracting congruent from incongruent interference trials which we do not report separately here.

Again, Bayes factor analyses provides evidence in favour of the null hypothesis of no difference between ego-depleted and control participants, $BF = 4.10$.

Thus even though performance during the ego-depletion task is markedly different between controls and ego-depleted participants and across texts, there is no difference in the subsequent inhibitory control task.

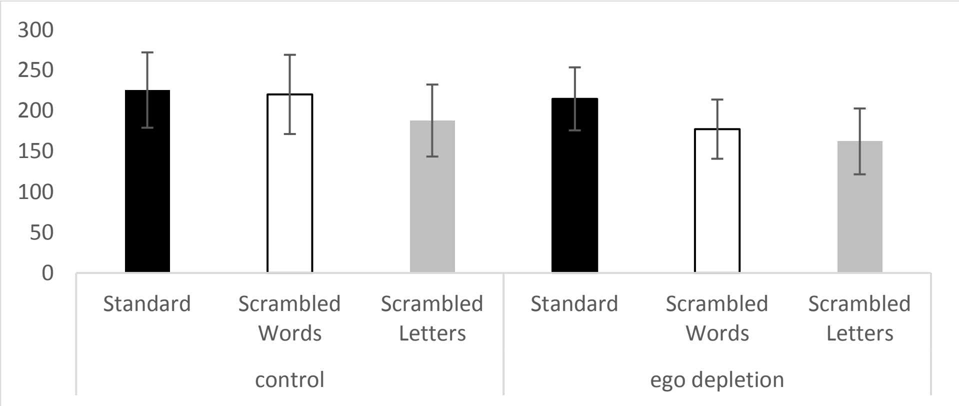

Figure 1. Mean Stroop inhibitory response time in milliseconds per condition and text. Error bars are 95% CI

Relation between ego-depletion and outcome task performance

To investigate which performance factors during the letter “e” task affect Stroop performance, Pearson correlation between the inhibitory control measure and condition, text, total marks set, proportion correct, and missing characters was conducted (Table 3). Condition (ego depleted versus control) correlated with performance on the letter “e” task whereas it did not with performance on the Stroop task. Thus, there is no sign of a relation between performance on the letter “e” task as a function of condition and subsequent Stroop performance.

Table 3. Correlations between condition, text manipulation, performance during the letter task and Stroop overall RT and inhibition RT.

	Condition	Text	RT overall	RT inhibtion	Total marks	Correct	Missing
Condition	---	-.04	.001	-.08	-.94***	-.48***	.45***
Text		---	-.02	-.12*	.04	.06	-.14*
RT overall			---	.35***	-.08	-.15**	.07
RT inhibtion				---	.04	-.08	-.06
Total marks					---	.39***	-.51***
Correct						---	-.32***
Missing							---

Experiment 2

In the second Experiment, the text remained constant (standard text from neuroscience article) while the complexity of the rules was manipulated.

Method

Participants

Overall 238 adults (191 females), *M* = 23 years, *SD* = 8 years, recruited via the xxxx University online participation system took part. Participants received course credit or financial reimbursement for participation.

Design

First, participants received the ego depletion task that contained rules of differing complexity (labelled 1-5), manipulated between participants. After that, all participants received the same computerized Stroop task as in Experiment 1. The whole experiment lasted around 30 minutes.

Materials and Procedure

As in Experiment 1 participants in the *control* rule (*N* = 39) were instructed to cross off all occurrences of the letter “e” (rule complexity level = 1). Participants in the “o not next to another vowel” rule (*N* = 38) were instructed to cross off all occurrences of the letter “o” except those that were next to another vowel (complexity level = 2). Participants in the “e more than one letter away from another vowel” rule (*N* = 48) were instructed to cross off any “e” that was more than one letter away from another vowel (complexity level = 3). Note, that this rule is essentially the same as the ego depletion rule but only using the second rule, thus, being less complex in wording. Participants in the standard *ego-depletion* rule (*N* = 53) crossed off an “e” that is not adjacent to another vowel or one extra letter away from another vowel (complexity level = 4). Participants in the “o not next to another vowel and e at least more than one letter away from another vowel” rule (*N* = 60) were instructed to cross off any o that is not next to another vowel plus any e that is at least more than one letter away from another vowel (complexity level = 5) (Table 4). Complexity level is defined by i) number of rules, and ii) frequency of letters in the text that can be crossed (i.e., an “o” is less frequent than an “e”. Participant numbers are uneven in each rule group as in the more complex rules (e.g., the latter one) many participants appeared to fail to follow instructions. However, this was analysed systematically below.

As in the previous experiment, participants were given 8 example words to understand the rules and had the rules written in front of them on a piece of paper. The task was performed on a typewritten sheet of paper with the same standard text as in Experiment 1 and participants were again told that they had 8 minutes to cross off the letters and subsequently performed the Stroop task.

Table 4. Overview of rules used in ego depletion task between participants.

Rule complexity	Rule	Example cross off	Example do not cross off
1	Control: all e‘s	-	-
2	O’s that are not next to another vowel	O in cross, other, today, flower	O in boar, exception, float, geode
3	E’s that are more than one letter away from another vowel	E in other, fennel, error, text	E in friend, feather, energy, enact
4	Standard: E that is not next to another vowel or that is one extra letter away from another vowel	E in other, fennel, error, text	E in friend, feather, energy, enact
5	1) O that is not next to another vowel and	O in cross, other, folklore, blossom	O in boar, flour, toad, geode,
	2) E that is more than one letter away from another vowel	E in other, fennel, error, text	E in friend, feather, energy, enact

Results

Ego-depletion and Stroop task performances were analysed in several univariate ANOVAs with rule as between participants variable. The data are available here: https://doi.org/10.6084/m9.figshare.7575581

Ego-depletion task performance

Total marks set. Depending on the rule, there was a difference in how many characters were crossed off during 8 minutes, $F(4, 233)$, = 211.47, $p < .001$, $\eta_p^2 = .78$.

More characters were crossed off in the control condition 1 ($M = 337$) than in all other conditions (ps < .001). More characters were also crossed off in the “o not next to another vowel” condition (complexity = 2) ($M = 177$) than all remaining rules, ps < .001. More characters were crossed off in the “o-vowel, e more than one letter away from another vowel” (complexity = 5) ($M = 107$) than the two remaining conditions, ps < .017. The “e more than one letter away from another vowel” (complexity = 3) ($M = 66$) and ego depletion rule (complexity = 4) ($M = 77$) did not differ, $p = 1.00$.

Proportion correct. The proportion of correct characters crossed based on the total marks set also differed between rules, $F(4, 233)$, = 7.69, $p < .001$, $\eta_p^2 = .12$. Participants in the control rule 1 ($M = .94$) crossed more correct characters than participants in all other rules (rule 2: $M = .79$, rule 3 $M = .71$, rule 4 $M = .80$, rule 5 $M = .81$), ps < .016. There were no further differences between rule groups, ps > .07.

Proportion missing. The number of correct characters that were not crossed based on the total amount of characters crossed also differed between rules, $F(4, 233) = 20.48$, $p < .001$, $\eta_p^2 = .26$. There were fewer correct characters missing in the control group ($M = .19$) than all other rules, ps < .007, except under the “o not next to another vowel” rule 2 ($M = .28$) which did not differ, $p = 1.00$. There were also fewer missing characters in rule 2 than both rule 3 ($M = .78$, $p < .001$) and rule 5 ($M = .56$, $p = .002$). Both rules 4 ($M = 44$) and 5 had also fewer missing characters (ps < .008) than rule 3. There were no further differences, ps > .40.

Stroop performance

Accuracy was approaching ceiling in all three Stroop trials (interference: *M* = 95.0; word reading: *M* = 98.2; colour naming: *M* = 98.1), therefore, no further statistical analysis was conducted (Table 5).

Table 5. Mean accuracies (percentages) for the three types of Stroop trials as a function of rule complexity. Standard deviation is in parentheses.

	1	2	3	4	5
Interference	93.7 (15.1)	94.5 (16.8)	95.1 (13.0)	97.0 (7.76)	94.4 (14.7)
Colour naming	97.4 (8.85)	99.0 (3.09)	97.9 (8.03)	98.1 (7.64)	98.1 (8.85)
Word reading	97.4 (8.86)	99.6 (.64)	98.0 (8.03)	98.2 (7.66)	98.2 (7.19)
Overall	96.2 (9.68)	97.7 (5.86)	97.0 (8.64)	97.8 (7.64)	96.9 (8.2)

Examining overall Stroop response time of correct trials as a function of condition and text, there was no difference between rules (rule 1: *M* = 1004 ms, rule 2: *M* = 976 ms, rule 3: *M* = 1004 ms, rule 4: *M* = 978 ms, rule 5: *M* = 975 ms), $F(4, 233) = .41$, $p = .80$, $\eta_p^2 = .01$. The same results were obtained for the inhibitory component of the task (mean response times of both word reading and colour naming trials subtracted from the mean of interference trials), $F(4, 233) = .86$, $p = .49$, $\eta_p^2 = .02$ (Figure 2). The same results were obtained when the Stroop effect was calculated differently by subtracting congruent from incongruent interference trials which we do not report separately here.

To examine the basic ego-depletion effect further, the Bayes factor was calculated comparing both overall response time and inhibitory response time between participants in the standard ego-depletion rule and control participants. For both overall RT and inhibitory RT, the Bayes factor favoured the null hypothesis, *BF* = 2.55, *BF* = 5.85, respectively, meaning there is no difference between the groups.

Thus, even though performance during the ego-depletion task is markedly different between rule groups, there is no difference in the subsequent inhibitory control task.

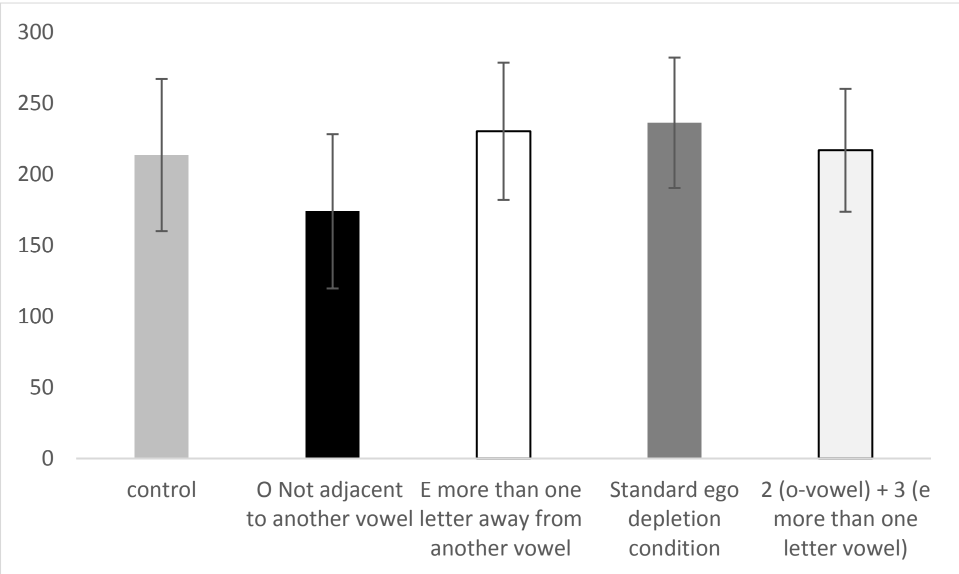

Figure 2. Mean Stroop inhibition response time in milliseconds per rule condition. Error bars are 95% CI

Relation between ego-depletion and outcome task performance

To investigate how performance factors during the different rule tasks relate to Stroop performance, a correlation between the inhibitory control measure and rule, total marks set, proportion correct, and missing characters was conducted (Table 6). The type of rule participants received correlated with performance on the ego-depletion task whereas it did not with performance on the Stroop task. Thus, there is no sign of a relation between performance on the ego-depletion task as a function of rule and subsequent Stroop performance.

Table 6. Correlations between rule condition, performance during the letter task, and Stroop overall RT and inhibition RT

	Rule	RT overall	RT inhibition	Total marks	Correct	Missing
Rule	---	.02	.06	.32***	.22***	-.07
RT overall		---	.60***	-.08	.01	.02
RT inhibition			---	-.11	-.02	.04
Total marks				---	.19**	-.46***
Correct					---	-.25***
Missing						---

Discussion

The current aim was to examine processes underlying ego-depletion by quantifying the level of ego-depletion and its subsequent effects on inhibitory control. In Experiment 1, manipulating the text from being semantically meaningful to non-meaningful showed that ego-depleted participants crossed off fewer letters *per se*, and made more mistakes and omissions than control participants. The type of text also increased difficulty such that more mistakes were made when the text had scrambled characters than scrambled words. This supports evidence suggesting that letter detection becomes worse in meaningful words because they are often identified as a whole unit (Healy, 1994). However, it should also be noted that neither differed from meaningful text. One possibility is that context “turned on” letter detection. Specifically, semantically meaningful context primes word recognition which in turn facilitates target letter recognition (Rumelhart & McClelland, 1982) and that might have counteracted the target letter omission error in standard text.

However, none of these performance differences had any relation with subsequent Stroop performance and the general ego-depletion effect was not found. In Experiment 2 the complexity of the ego-depletion rule was manipulated while the text remained constant. Again, the overall pattern was the more complex the rule, the worse the performance such that participants crossed off fewer characters, made more mistakes and missed more correct characters. Yet again, ego-depletion performance was unrelated to subsequent Stroop performance and there was no evidence of the typical ego-depletion effect. Power analyses (Gpower, Faul, Erdfelder, Lang, & Buchner, 2007) showed that 210 participants were needed in both experiments to detect differences between participants who received the standard ego-depletion instructions and control participants in Stroop inhibitory control performance to detect medium sized effects, $f = .25$. Thus, we clearly had sufficient power to detect differences.

Overall, the lack of an effect of the letter cancellation task partially fits recent findings, for example, the preregistered replication study (Hagger et al., 2016), Bayesian analyses of the effect (Etherton et al., 2018) and behavioural results with a large sample size that found a very small effect (Wimmer, Stirk, & Hancock, 2017). However, current results can only be directly compared to the latter study where the same ego-depletion task and Stroop task as outcome task pairing was used. Comparison to other studies is difficult as the letter cancellation task was paired with a different outcome task. For example, Wimmer et al. (2017) found medium effects in that participants are less likely to perceive interpretations of ambiguous figures after having completed the letter cancellation task whereas the effect was small for Stroop performance. Moreover, other studies have found typical ego-depletion effects for the letter cancellation task paired with different outcome measures (Baumeister et al., 1998; DeWall, Baumeister, Gailliot, & Maner, 2008; Stripada et al., 2014; Wan & Sternthal, 2008), making comparison between studies difficult, especially as there is no conceptualization how the tasks tap into self-control (Lurquin & Miyake, 2017).

Additionally, one of the main critiques of Haggart et al.'s (2016) replication study is that they did not use a habituation phase where participants first crossed off all e's and then were given the ego-depletion rules, thus, requiring self-control (Baumeister & Vohs, 2016). However, nowhere in earlier studies nor in the original study (Baumeister et al., 1998) was such a habituation procedure reported and the effect has been shown without the habituation procedure (Baumeister et al., 1998; Wan & Sternthal, 2008) nor is there experimental evidence that having a habituation procedure makes the task more difficult than without (e.g., Arber et al., 2017). Thus, this argument seems a post-hoc application to the data.

Where do the current findings leave us with regard to the ego-depletion debate? First, there is now quite substantial evidence that when the letter e task is paired with a Stroop task as outcome measure, then the effect is not shown or is very small, as shown with large sample sizes (current study Experiment 1: $N = 102$, Experiment 2: $N = 99$, only including the standard ego-depletion and control scenarios, and Wimmer et al., 2017: $N = 214$). The Stroop task is one of the most established measures of inhibitory control and there was no effect of condition on Stroop performance, suggesting that letter cancellation does not cause ego depletion. There was also no correlation between ego-depletion task performance and the Stroop task.

Additionally, performance on the letter cancellation task was affected by both text manipulation and complexity, but it is unclear whether participants who performed less correctly in the cancellation task were more or less depleted. They might perform worse because of being more depleted or just not concentrating and therefore less depleted than others who worked harder and performed better. However, task difficulty should enhance the chances of finding positive ego-depletion effects (Hagger et al., 2010) and we have clear evidence that our manipulations systematically altered task difficulty. Thus, altogether, findings raise doubts about the suitability of the letter cancellation task for ego-depletion research if one subscribes to the view that ego-depletion is a lack of self-control which is akin to lack of inhibitory control.

A plausible alternative interpretation of the findings is that ego-depletion as such does not exist but is confined to a narrow subset of manipulation task-outcome task pairings. Recent carefully designed studies have consistently failed to replicate the ego-depletion effect, while meta-analyses have also found no effect (Carter, et al., 2015; Etherton, et al., 2018; Hagger et al., 2016; Lurquin et al., 2016).

It is also noteworthy, that the reported ego-depletion effect is in the opposite direction to processes underlying other subsequent task paradigms in Cognitive Psychology such as sequential modulation. Here, within a single task, when presented with trials that pose response conflict, subsequent trial performance requiring inhibitory control is enhanced due to activated cognitive control or priming (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Mayr, Awh, & Laurey, 2003). For example, responding to incongruent Flanker task trials reduces response time in Stroop interference trials (Fernandez-Duque & Knight, 2008), thus showing the opposite effect from what is assumed to occur during ego-depletion. The idea of reversed ego-depletion has recently been put forward where the more difficult the task, the better performance was afterwards, when in line with cultural norms (Savani & Job, 2017). However, as ego-depletion as such is not conceptualized it is unclear which processes underlie reverse ego-depletion and whether these might tap into similar processes underlying sequential modulation.

In sum, the current research suggests that the letter cancellation task is not a suitable index for examining the effects of ego-depletion if one conceptualizes self-regulation as an index of inhibitory control. We suggest future research should continue to systematically vary ego-depletion task difficulty to quantify the level of ego-depletion needed to have subsequent effects on task performance and to test whether the effect exists to begin with.

References

Alós-Ferrer, C., Hügelschäfer, S., & Li, J. (2015). Self-control depletion and decision making. *Journal of Neuroscience, Psychology and Economics, 8*, 203-216. doi: 10.1037/npe0000047

Arber, M. M., Ireland, M. J., Feger, R., Marrington, J., Tehan, J., & Tehan, G. (2017). Ego depletion in real-time: An examination of the sequential-task paradigm. *Frontiers in Psychology, 8*. doi: 10.3389/fpsyg.2017.01672

Baumeister, R. F. (2014). Self-regulation, ego depletion, and inhibition. *Neuropsychologia, 65,* 313 -319. doi: 10.1016/j.neuropsychologia.2014.08.012

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., Tice, D. M., & Vohs, K. D. (2018). The strength model of self-regulation: Conclusions from the second decade of willpower research. *Perspectives on Psychological Science, 13*, 141-145. doi:10.1177%2F1745691617716946

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

Botvinick, M. M, Braver, T. S., Barch, D. M., & Carter, C. S. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108, 624-652*. doi:10.1037/0033-295X.108.3.624

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

DeWall, C. N., Baumeister, R. F., Gailliot, M. T., & Maner, J. K. (2008). Depeltion makes the heart grow less helpful: Helping as a function of self-regulatory energy genetic relatedness. *Personality and Social Psychology Bulletin, 34*, 1653-1662. doi: 10.1177%2F0146167208323981

Eaton, J. W., Bateman, D., Hauberg, S., & Wehbring, R. (2017). GNU Octave version 4.4.1 manual: a high-level interactive language for numerical computations. URL https://www.gnu.org/software/octave/doc/v4.4.1/

Etherton, J. L., Randall, O., Stephenson, K., Grace, M., Jones, C., & De Nadai, A. S. (2018). Bayesian analysis of multimethod ego-depletion studies favours the null hypothesis. *British Journal of Social Psychology, 57*, 367-385. doi:10.1111/bjso.12236

Evans, D. R., Boggero, I. A., & Serstrom, S. C. (2016). The nature of self-regulatory fatigue and "ego depletion". *Personality and Social Psychology Review, 4*, 291-310. doi:10.1177%2F1088868315597841

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*, 175-191. doi:10.3758/BF03193146

Fernandez-Duque, D., & Knight, M. B. (2008). Cognitive control: Dynamic, sustained, and voluntary influences. *Journal of Experimental Psychology: Human Perception and Performance, 34*, 340-355. doi:10.1037/0096-1523.34.2.340

Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H. et al. (2016). A multilab preregistered replication of the ego-depletion effect. *Perspectives on Psychological Science, 11*, 546-573. doi:10.1177%2F1745691616652873

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin, 136*, 495-525. doi:10.1037/a0019486

Inzlicht, M., & Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science, 7*, 450-463. doi:10.1177%2F1745691612454134

Inzlicht, M., Schmeichel, B. J., & Mcrae, N. C. (2014). Why self-control seems (but may not be) limited. *Trends in Cognitive Sciences, 18*, 127-133. doi:10.1016/j.tics.2013.12.009

Lurquin, J. H., & Miyake, A. (2017). Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2017.00568

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience, 6*, 450-452. doi:10.1038/nn1051

Muraven, M., & Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin, 29*, 894-906. doi:10.1177/0146167203029007008

Myers, L., Downie, S., Taylor, G., Marrington, J., Tehan, G., & Ireland, M. J. (2018). Understanding performance decrements in a letter-cancelling task: Overcoming habits or inhibition of reading. *Frontiers in Psychology, 9*. doi: 10.3389/fpsyg.2018.00711

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*, 589-607.

Stripada, 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%2F0956797614526415

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662. doi: 10.1037/h0054651

Vonash, A. J., Vohs, K. D., Pocheptsova Gosh, A., & Baumeister, R. F. (2017). Ego depletion induces mental passivity: Behavioral effects beyond impulse control. *Motivation Science, 3*, 321-336. doi:10.1037/mot0000058

Wan, E. W., & Sternthal, B. (2008). Regulating the effects of depletion through monitoring. *Personality and Social Psychology Bulletin, 34,* 32-46. doi: 10.1177%2F0146167207306756

Wimmer, M. C., Stirk, S., & Hancock, P. J. B. (2017). Ego depletion in visual perception: ego-depleted viewers experience less ambiguous figure reversal. *Psychonomic Bulletin and Review, 24*, 1620-1626. doi: 10.3758/s13423-017-1247-2