Study 4 Overview: Stroop as DV

In our previous high-powered pre-registered ego-depletion study using standardized tests as the dependent variable (Lin et al., 2016c, OSF osf.io/375zg), we failed to replicate previous ego depletion effects. This was despite the large differences in self-report ratings of mental demand and effort between the control and depleting conditions, d [95 %CI] = -3.89 [-4.37 , -3.42]. In this study, we will use the same modified symbol counter task to induce depletion but will instead use the Stroop task as the dependent variable because it is often used in ego-depletion studies (e.g., Carter et al., 2015) and is a reliable measure of cognitive control.

Experimental conditions

There are two conditions in this study: control and depletion. After completing the control (watching a 5-minute wildlife video) or depleting task (modified symbol counter task), participants will complete questions that assess subjective experience (i.e., mental demand, effort, frustration, boredom; see Lin et al., 2016a, 2016b, 2016c). We will also measure mental fatigue using one item: “I’m mentally fatigued now.”

Design

In our previous pre-registered studies (Lin et al., 2016a, 2016b, 2016c), we deviated from previous ego-depletion work by using a pre-post design (i.e., dependent measures were measured twice, before and after experimental manipulation, with first measure as the covariate), which we believed could increase statistical power. Although statistically sound, a pre-post design might have unwittingly led to subtle changes in motivation and behavior, which might have negated depletion effects.

In the present study, we will recruit undergraduate psychology students (our previous studies used Amazon Mechanical Turk workers). Each participant will complete both the depleting and control versions of the task in our laboratory, but on separate days. This within-subjects design will give us greater statistical power. Given that psychology undergraduates are usually very familiar with the Stroop task, we do not believe that completing the task twice, on different days, could affect depletion effects. Participants will be pseudorandomly assigned to either the control or depleting condition on the first

day based on their allocated participant number (i.e., odd-numbered participants will be assigned to the control condition; even-numbered participants will be assigned to the depleting condition). To minimize demand characteristics, the two sessions/days will be separated by at least a week.

Sample and power

Based on the meta-analytic effect sizes for the Stroop task (d = 0.24; Carter et al., 2015), a power analysis indicates that 166 participants will be sufficient to achieve .80 power (α = .05) for our within-subjects design (assuming correlation between repeated measures: .40). Our target sample size is therefore 166.

Although classical frequentist testing forbids optional stopping, Bayesian hypothesis testing allows both optional stopping and multiple testing. We will use sequential Bayes factor and will stop data collection if the Bayes factor favoring the experimental hypothesis is 10 or greater (see statistical analyses section below). However, if the Bayes factor fails to reach 10, we will stop data collection when we reach our target sample size of 166.

Dependent measure: Stroop error interference

There are many possible Stroop outcome measures, but we are expecting depletion effects on Stroop interference: % incongruent error trials minus % congruent error trials. Nevertheless, we will look at other Stroop performance indices (e.g., reaction time and accuracy on incongruent trials), but these analyses will be exploratory. These are the parameters for our Stroop task:

- Three colors (red, blue, yellow)
- 6 practice trials
- 180 actual trials (120 congruent, 60 incongruent) presented in 1 block
- If participants respond in 2000 ms, the total trial duration (response time + intertrial interval) will always be 2000 ms. But if they fail to respond in 2000 ms, a message will be shown at the 5000th ms to remind them to respond faster ('Too slow! Response required.') and this message will be shown until a response has been registered.

Exclusion criteria

Modified symbol counter task (depleting task; see Lin et al., 2016b, 2016c): Participants who get fewer than 35 trials correct (out of 60) will be excluded.

Stroop task

- Trials with RT < 200 ms
- Trials with RT > 3 median absolute deviations (MADs; Leys et al., 2013) from individual's overall RT
- Individuals whose total number of errors on congruent trials > 3 MADs from the overall median across both experimental conditions
- Stroop interference (% incongruent error trials minus % congruent error trials) > 3 MADs, within each experimental condition

Data quality: At the end of the experiment, participants will be asked to rate (1 to 7) whether they think they have properly completed the tasks and have provided good data. They will be told that they will be compensated regardless of their response to this question. Anyone who provides a rating of 2 or lower will be excluded from analyses.

Individual difference measures to test for moderation

- Implicit willpower beliefs (6 items; Jobs et al., 2010)
- Big Five (conscientiousness facet; 9 items; John & Srivastava, 1999)
- Action orientation measured using the Demand-Related Action Orientation subscale of the Action Control Scale (12 items; see Gropel et al., 2014)
- Brief self-control scale (13 items; Tangney et al., 2004)
- SRSQ self-control scale (9 items; Imhoff et al., 2014)
- Age

All the questionnaires above will be administered at the end of the second session/day.

Hypotheses

After performing the depleting (vs. control) task, participants should perform worse (i.e., increased Stroop interference) on the Stroop task.

Statistical analyses

When the total sample size *n* is smaller than 166, we run use sequential Bayes factor (Schönbrodt et al., 2015, 2016) during data collection to assess the strength of evidence in favor of our experimental hypothesis. Unlike classical frequentist testing, the Bayesian approach allows for unlimited multiple testing, even after each participant. We will stop data collection once the Bayes factor is 10 or greater in favor of the experimental hypothesis any point during data collection. Thus, our final sample size might be smaller than 166. We will only use classical frequentist methods—dependent t-test and multi-level model (condition nested within participant)—once our Bayes factor is 10 or greater. That is, we will only use classical frequentist methods once we have stopped data collection.

If Bayes factor fails to reach 10 during data collection, we will stop data collection when the sample size reaches 166. Then, we will test our primary hypothesis using a dependent t-test and multi-level model.

We will supplement all our classical frequentist hypothesis testing with Bayes factor (Cauchy prior $r = 1$).

References

Gröpel, P., Baumeister, R. F., & Beckmann, J. (2014). Action versus state orientation and self-control performance after depletion. Personality and Social Psychology Bulletin, 40(4), 476-487. doi:10.1177/0146167213516636

Imhoff, R., Schmidt, A. F., & Gerstenberg, F. (2014). Exploring the interplay of trait self-control and ego depletion: Empirical evidence for ironic effects: Ironic effects of self-control. Eur J Pers, 28(5), 413-424. doi:10.1002/per.1899

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. doi:10.1177/0956797610384745

John, O. P., & Srivastava, S. (1999). The Big-Five trait taxonomy: History,

measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), Handbook of Personality: Theory and Research (Vol. 2, pp. 102-138). New York: Guildford Press.

Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology, 49(4), 764-766. doi:10.1016/j.jesp.2013.03.013

Lin, H., Inzlicht, M., Saunders, B., & Friese, M. (2016a, June 7). Pre-registration plan. Retrieved from osf.io/bdek5

Lin, H., Inzlicht, M., Saunders, B., & Friese, M. (2016b, June 17). Replication of pre-registered study. Retrieved from osf.io/tfa3s

Lin, H., Inzlicht, M., Saunders, B., & Friese, M. (2016c, August 16). Study 3: Standardized tests as DV. Retrieved from osf.io/375zg

Schönbrodt, F. D., & Wagenmakers, E. J. (2016). Bayes factor design analysis: Planning for compelling evidence. Available at SSRN 2722435, 4(3), 187-201. doi: 10.1002/(ISSN)1539-1612

Schönbrodt, F. D., Wagenmakers, E. J., & Zehetleitner…, M. (2015). Sequential hypothesis testing with bayes factors: Efficiently testing mean differences. psycnetapaorg. doi:10.1037/met0000061

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self control predicts good adjustment, less pathology, better grades, and interpersonal success. Journal of Personality, 72(2), 271-324.