Misguided Effort with Elusive Implications

Roy F. Baumeister & Kathleen D. Vohs

In retrospect, the decision to use new, mostly untested procedures[1] for a large replication project was foolish. When planning the RRR on ego depletion, Hagger asked Baumeister for suggestions. Baumeister nominated several procedures that have been used in successful studies of ego depletion for years. But none of Baumeister's suggestions were allowable due to the RRR restrictions that it must be done with only computerized tasks that were culturally and linguistically neutral. Discussions were stalemated and we felt time pressure to come up with something. We learned of a new study by Sripada et al. that fit the requirements, and passed this along to the RRR. Since there were no viable other options, that method was chosen.

Apparently it matters how much we endorsed this method. To be clear: No one working in either of our laboratories has ever used this procedure in any study (neither the manipulation nor dependent measure). We still do not understand why reaction time variance is a measure of self-control failure (are people overriding some impulse to react with variable speed?), but the idea of "replication" requires that something like the task have been used at least once, and the Sripada et al. paper reported successful results in a major outlet. (Perhaps we should have still objected. But because there were no other viable options, objecting would have meant objecting to the entire RRR, which could have been interpreted as lack of trust in the effect.) Under the circumstances, we understood our approval to mean "Sure, go ahead" and not "yes, that's a definitive test of the phenomenon we've been studying all these years." Crucially, we thought the robustness of ego depletion effects would overcome any weaknesses in this new method. That was an unfortunate mistake, partly because the weaknesses seem more serious than we had understood.

The manipulation is a computerized version of what is called the e-crossing procedure. This procedure was originally created as a laboratory version of a common self-control task, namely breaking a habit. Self-regulation is typically understood as altering and overriding responses. The e-crossing task works because participants first establish a habit (of using a pencil to cross out every e on a page of text) and then must override these habitual responses when more complex rules are introduced. Self-

---

[1] To be precise, the procedure was very loosely based on a familiar one, but in this particular form was used only once by Sripada et al. (in press at the time). This procedure is quite different from what typical depletion studies have used. Indeed, aside from the RRR, we do not know of any other study using the Sripada method since its publication. Further, Sripada's commentary highlights how the RRR did not precisely copy his method. Hence, technically, it was a novel, untested procedure.

regulation is invoked when the participant sees an e and experiences the impulse to cross it off — and then must restrain that impulse. The Sripada and RRR studies skipped the initial key step of establishing a habit. RRR participants simply pressed a button to indicate whether each word has an e that is not adjacent to another vowel. Without first instilling the habit, there is nothing to override. This may be a difficult cognitive judgment task, but no impulse is overridden, contrary to the nature of self-control tasks.

The RRR says that skipping the initial habit-forming step was justified because other tasks in the literature have done the same, such as a manipulation in which participants are instructed to write a story with words that do not contain the letters A or N (originally by Schmeichel, 2007). Yet that task is depleting precisely because there most certainly is a very strong habit. An English speaker has spent years writing sentences using all letters of the alphabet, including A and N. This misunderstanding highlights what may be a problem in the field as a whole in its current focus on replication: It is misguided to focus merely on the simple structure of procedures while disregarding the underlying psychological processes. Scientific hypotheses concern psychological processes, not laboratory procedures.

Self-report data from the RRR suggest that the task does not involve self-regulation. Manipulation checks are difficult to obtain with ego depletion, because people cannot usually report on subjective changes indicative of having expended resources in self-regulation. The closest to a reliable measure is self-reported fatigue; negative mood may increase slightly (Hagger, Wood, Stiff, & Chatzisarantis, 2010). Self-report data indicated that RRR participants found the task extremely frustrating but not fatiguing — unlike the usual pattern in ego depletion.

One question going forward is how to create replication studies that are not constrained to computerized methods stripped of contextual factors. The admirable ideal that all meaningful psychological phenomena can be operationalized as typing on computer keyboards should perhaps be up for debate (Baumeister, Vohs, & Funder, 2007). Computer-administered measures of executive functioning apparently relate poorly to self-control phenomena (Duckworth & Kern, 2016), though that was not known when the RRR started. Purely cognitive tasks may be an ineffective method for studying ego depletion (Inzlicht, Gervais, & Berkman, 2016), and researchers may do better by focusing on impulsive, emotional, behavioral, and brain effects. Theories might consider the possibility that ego depletion does not affect cognitive processes directly but rather disrupts their connection to other brain regions (Kelley, Wagner, & Heatherton, 2015).

In all this and the so-called replication crisis generally, two different questions often are being confused. One concerns the generality of causal principles, and the other the reliable effectiveness of particular lab procedures. If an experiment fails to manipulate the independent variable, it does not test the hypothesis. Signs indicate the RRR was plagued by manipulation failure — and therefore did not test ego depletion.

For two decades we have conducted studies of ego depletion carefully and honestly, following the field's best practices, and we find the effect over and over (as have many others in fields as far-ranging as finance to health to sports, both in the lab

and large-scale field studies). There is too much evidence to dismiss based on the RRR, which after all is ultimately a single study — especially if the manipulation failed to create ego depletion.

Clearly, though, this debacle shifts the burden of proof onto those of us who believe ego depletion effects are genuine. We will organize a pre-registered, multi-site replication project next year, using well-tested procedures (ones that actually involve self-regulation). We herewith preregister the hypothesis the depleted participants will perform worse on subsequent, ostensibly unrelated self-regulation tests than non-depleted participants, as a great many other studies have found.

References

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*, 396-403. doi: 10.1111/j.1745-6916.2007.00051.x

Duckworth, A.L, & Kern, M.L. (2016). A meta-analysis of the convergent validity of self-control measures. Manuscript submitted for publication, University of Pennsylvania.

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

Inzlicht, M., Gervais, W.M., & Berkman, E.T. (2016). Bias-correction techniques alone cannot determine whether ego depletion is different from zero: Comment on Carter, Kofler, Forster, and McCullough, 2015. Manuscript submitted for publication.

Kelley, W.M., Wagner, D.D., & Heatherton, T.F. (2015). In search of a human self-regulation system. *Annual Review of Neuroscience,* 38, pp. 389-411. doi: 10.1146/annurev-neuro-071013-014243.

Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology, 136*, 241-255. doi:10.1037/0096-3445.136.2.241