Manuscript ID: PSCI-15-1916-R4

IN PRESS: PSYCHOLOGICAL SCIENCE

(May 2016)

Title: The bitter truth about sugar and willpower: The limited evidential value of the glucose model of ego depletion

Authors: Miguel Vadillo[1], Natalie Gold [2], Magda Osman[3]

Affiliations:

[1] King's College London, Primary Care and Public Health Sciences, Guy's Campus, London SE1 1UL, UK

[2] King's College London, Department of Philosophy, Philosophy Building, London WC2R2LS, UK

[3] Queen Mary University of London, Biological and Experimental Psychology, School of Biological and Chemical Sciences, Mile End Road, London E1 4NS, UK

Correspondence author: Magda Osman

Keywords: Ego depletion, Glucose, Meta-analysis, P-curve, Self-control

Word count (Introduction + Discussion + Author Note): 920 + 846 + 81

Abstract

Ego depletion is the idea that willpower draws on a limited mental resource, so that engaging in an act of self-control impairs self-control in subsequent tasks. To present ego-depletion as more than a convenient metaphor, Gailliot et al. (2007) have proposed that glucose is the limited resource that becomes depleted with self-control. However, there have been theoretical challenges to the glucose mechanism and the experiments that test it have found mixed results. We use a new meta-analytic tool, *p*-curve, to examine the reliability of the evidence from these experiments. We find that the effect sizes reported in this literature are possibly influenced by publication or reporting bias and that, even within studies yielding significant results, the evidential value of this research is weak. In light of these results, and pending further evidence, researchers and policy makers should refrain from drawing any conclusions about the role of glucose in self-control.

(Word count: 149)

According to the law of conservation of energy, the total amount of energy of an isolated system can never increase. In the domain of psychology, the idea that energy is a limited resource originated with Freud (1933/1961). Energy models have been little used in psychology since Freud, though, with the rare exception of the *ego depletion* model developed by Baumeister, Bratslavsky, Muraven, and Tice (1998). According to this research, "the self's acts of volition draw on some limited resource, akin to strength or energy and that, therefore, one act of volition will have a detrimental impact on subsequent volition" (Baumeister et al., 1998, p. 1252). The implications of research on ego depletion are considerable. It has been claimed that reliably exerting self-control, both in terms of actively doing something good and avoiding the temptation to act on bad impulses, can greatly reduce many of the major ills that effect our society and our personal lives, such as "crime, violence unwanted pregnancy, drug addiction, venereal diseases, bankruptcy, and premature deaths" (Baumeister, Muraven, & Tice, 2000, pp. 130). Not surprisingly, the work of Baumeister et al. (1998) has impacted a number of disciplines, including advertising, behavioral economics, business, consumerism, law, management, marketing, and medicine. In fact, it is fair to say that the seminal paper by Baumeister et al. (1998) has become a classic, with more than 1,250 citations in the *Web of Science* at the time of writing the present article.

When ego depletion was first proposed, the idea of a limited resource was a convenient metaphor. Given how fundamental exerting self-control is thought to be, it is important to establish the energy source that is depleted and to provide a mechanism by which ego depletion occurs. The most popular candidate that we find in the literature involves glucose. Gailliot et al. (2007) presented nine studies supporting three main findings: (1) Blood glucose levels are reduced after performing a self-control task but not after performing a comparative cognitive task that does not require self-control; (2) low levels of blood glucose after a first self-control task predict behavioral deficits on a second self-control task; and (3) participants whose glucose levels are restored after a self-control task, by ingesting a glucose drink, perform better on a subsequent task than participants who are given a diet drink in between. They concluded that self-control depletes blood glucose leading to decreased self-control on subsequent tasks, and restoring glucose levels replenishes the ability to exert self-control.

While Gailliot et al.'s (2007) conclusions have been extraordinarily influential, their glucose hypothesis remains controversial. The mechanism they propose has been challenged and the reliability of their results has been disputed. Kurzban (2010) argues that the glucose mechanism as presented by Gailliot et al. (2007) is biologically implausible. The mechanism invokes the idea that self-control tasks deplete glucose because of energy consumption by the *brain*, but the supporting evidence only shows changes in *blood* glucose levels. Kurzban cites evidence that the sort of self-

control tasks used in the literature have little effect on brain metabolism and that changes in blood glucose are unlikely to reflect blood glucose uptake by the brain.

There are also concerns about the empirical evidence for the glucose mechanism. For instance, Schimmack (2012) showed that the number of significant results reported by Gailliot et al. (2007) is too large, given their average power. In other words, these results are likely to be influenced by publication bias and/or *p*-hacking (see Francis, 2012; Simmons, Nelson, & Simonsohn, 2011). His concerns are supported by a reanalysis of Gailliot et al.'s data, which finds that self-control does not decrease blood glucose levels (Kurzban, 2010), and by recent failures to replicate the effect of glucose on self-control (Job et al., 2013; Kelly et al., 2015; Lange & Eggert, 2014; Lange et al., 2014).

Generally there are three broad categories of studies that examine the effect of glucose on self-control: "correlational studies", and two types of intervention studies, "glucose ingestion" and "glucose rinsing". Correlational studies measure before/after effects on blood glucose levels when self-control is exerted (Dvorak & Simons, 2009; Galliot, 2012; Gailliot et al., 2007; Molden et al., 2012). Glucose ingestion studies manipulate blood glucose levels, leaving enough time between ingestion and the control task for glucose to be absorbed into the blood stream (Birnie et al., 2015; Denson et al., 2010; DeWall et al., 2008; Dickinson et al., 2014; Gailliot et al., 2007; Gailliot et al., 2009; Howard & Marczinski, 2010; Job et al., 2013; Kelly et al., 2015; Lange & Eggert, 2014; Lange et al., 2014; Masicampo & Baumeister, 2008; Wang & Dvorak, 2010). While the findings provide mixed support for the glucose hypothesis, the methods used imply a mechanism that is consistent with Gailliot et al.'s proposals. A third set of studies ("glucose-rinsing studies") examine the impact of simply rinsing one's mouth with glucose prior to exerting self-control (Hagger & Chatzisarantis, 2013; Lange & Eggert, 2014; Molden et al., 2012; Sanders et al. 2012). These studies suggest that the signal of glucose from the mouth to the brain is sufficient to neutralize the ego-depletion effect. This mechanism is consistent with the results of the ingestion manipulations, but it suggests that the effect does not depend on a metabolic explanation.

To help settle the growing concerns in the academic community regarding the reliability of the glucose mechanism, which in turn has implications for the ego-depletion hypothesis that it underpins, we sought to use a new meta-analytic tool, *p*-curve, to investigate the presence of publication and reporting biases in this literature.

Method

Literature search strategy

We looked for studies supporting the idea that sugar consumption is related to ego-depletion and self-regulation. Specifically, we took into account any study exploring the hypothesis that glucose ingestion or rinsing improves performance in a laboratory task demanding self-regulation (e.g., overcoming an impulse, inhibiting an aggressive reaction, or controlling a cognitive process) or ameliorates the effect of an ego-depleting experience on these laboratory self-regulation tasks. We also included studies testing whether performance in laboratory tasks is correlated with pre- or post-testing sugar levels, again, specifically concerning tasks that explicitly require self-regulation. Studies in which participants were not asked to drink sugar, but simply to rinse their mouths with sugar, were also included in the present analyses, because this literature also supports the idea that sugar consumption improves self-regulation (even if it questions the specific hypothesis that this is achieved through metabolic processes).

Given these criteria, we excluded experiments showing a relation between sugar consumption and cognitive processes that, prima facie, do not seem to pose demands on self-regulation like, for instance, short-term memory or general cognition function (e.g., Carter & McCullough, 2013; Owen, Scholey, Finnegan, Hu, & Sünram-Lea, 2012). In addition, based on the conditions we set, studies were also excluded that investigated the correlation between general glucose levels (or regular glucose ingestion) and self-regulated behavior across many days in naturalistic settings. This includes, for instance, studies on the relationship between glucose ingestion and smoking cessation or diabetes and different psychological processes (see Gailliot & Baumeister, 2007). These studies rely on measures that differ substantially from the dependent variables gathered in laboratory-based ego-depletion tasks and the lack of experimental control makes the results amenable to alternative explanations that bear little or no relation to ego depletion and self-regulation.

We began our search by inspecting a small set of studies that had included an exhaustive literature review. These included Hagger et al.'s (2010) meta-analysis on the general ego-depletion literature, a study by Job et al. (2013) exploring inter-individual differences in the impact of glucose on self-control, and Lange and Eggert's (2014) recent attempt to replicate the effect of sugar consumption or rinsing on ego-depletion. To make sure that we included all relevant studies, we then conducted a systematic search in the *Web of Science* and *Google Scholar* for the term "glucose" and "ego depletion" or "self-control" or "self-regulation". This allowed us to detect 18 articles with one or more eligible studies. All these studies are listed in Table 1 and are also marked with asterisks in the reference list. Furthermore, we found out that one of our selected studies (Masicampo & Baumeister, 2008) had been included in the famous project on the reproducibility of psychological

science (Open Science Collaboration, 2015). This replication was also included in our analyses yielding a total of 19 articles.

P-curve

Our assessment of the reliability of this set of studies was based on *p*-curve (Simonsohn, Nelson, & Simmons, 2014), a recently designed meta-analytic tool that explores various biases in the published literature using only the distribution of significant *p*-values. Imagine that a set of studies is exploring an effect that does not exist. Occasionally, these studies will observe a significant result just by chance (i.e., a *p*-value lower than .05). In this scenario, all *p*-values will be equally likely: 5% of studies will have *p*-values lower than .05, 4% of studies will have *p*-values lower than .04, and so on. Consequently, the *p*-values of a set of studies exploring a non-existent effect should typically follow a flat distribution. Interestingly, this is not the case if the studies are exploring true effects. In this case, significant *p*-values should follow a right-skewed distribution, where small *p*-values (e.g., $p < .01$) should be more prevalent than larger *p*-values (e.g., $.04 < p < .05$). As explained by Simonsohn et al. (2014), this can be easily understood if one imagines an experimenter exploring a very large effect with a large sample of participants. Most likely, he/she will observe a very low *p*-value. Experiments with smaller effect sizes and smaller samples are simply less extreme versions of this ideal scenario. Even for low powered studies, the distribution of *p*-values should be right skewed. This implies that, in principle, one can know whether a set of experiments is exploring true effects or null effects simply by checking whether their *p*-values follows a right-skewed distribution or a rather flat distribution. An interesting feature of this approach is that it focuses exclusively on significant *p*-values (i.e., studies where $p < .05$) and, consequently, its results are unaffected by publication bias.

An online application designed by Simonsohn et al. (available at http://www.p-curve.com) allows researchers to test whether an observed distribution of *p*-values is significantly right-skewed or, alternatively, whether the distribution of *p*-values is suspiciously flat, which could suggest that the significant results are false positives. A simple way to test whether the distribution of *p*-values is significantly right-skewed is to compare the number of significant *p*-values lower than .025 with the number of p-values between .025 and .05 using a binomial test. A potential shortcoming of this approach is that this binomial test gives the same weight to exceptionally small *p*-values (e.g., .00001) as to *p*-values barely smaller than .025 (e.g., .024). To overcome this limitation, the latest versions of *p*-curve not only conduct a binomial test, but also an alternative analysis, known as continuous test, that is sensitive to the exact *p*-values.

If the distribution of *p*-values is not significantly right-skewed, this might mean that the studies lack any 'evidential value' or, in other words, that the significant results can be false positives. However, failure to find a significant right-skewed distribution might also be due to the lack of statistical power (e.g., if a very small number of studies is included in the analysis). As a means to test whether the distribution of *p*-values is too flat, Simonsohn et al. (2014) suggested testing whether the *p*-curve is flatter than the theoretical distribution that one would observe in a set of studies with 33% statistical power. If the *p*-curve is significantly flatter than this very flat standard, a common conclusion is that the set of studies might lack evidential value and that they might be the product of publication bias, selective reporting, or *p*-hacking.

Selection of statistical contrasts

We selected the key statistical contrasts of each study following the guidelines offered by Simonsohn et al. (2014). In the case of correlational studies or experiments with just two groups, we registered the target correlation coefficient, the statistic testing the regression slope, or the statistic testing the difference of means. In complex factorial designs, if researchers expected to find an ego-depletion effect disappearing in a specific condition, then we registered the statistic testing the interaction. In contrast, if they were expecting to find a complete cross-over interaction, we registered the two simple effects. A total of 38 statistical contrasts were included in the analyses. Following the recommendations of Simonsohn et al. (2014), when two statistics were equally valid, we used one of them in the main analysis and the other one in a robustness test. In most cases (4/5), we adopted the general rule of selecting the first one appearing in the text for the main analysis and the second one for the robustness test. However, in one occasion (Birnie et al., 2015) we broke this rule because the conclusions of the authors relied more heavily on one of the statistics. In this particular case, we selected this specific statistic for the main analysis and the other one for the robustness test. We did not find any study where three or more statistics were equally valid for the *p*-curve analysis. A *p*-curve disclosure table with all the selected statistics and the justification for our choice is available at http://bit.ly/1PoHFby.

Results

The list of studies, together with the key statistical contrasts of each study, are shown in Table 1. Figure 1 plots the frequencies of each range of *p*-values among the studies included in the present analysis. As can be seen, the main *p*-curve does not fit with the right-skewed distribution that one would expect if these studies were exploring a true effect. The only 'positive' feature of the

distribution is that there are no significant results immediately below .05. However, *p*-values in the [0, .04] interval show, if anything, a left-skewed distribution. Not surprisingly, the statistical contrast testing the right-skewness of the *p*-curve was non-significant, (p = .575 for the binomial test comparing the proportion of $p < .025$ and $.025 < p < .05$, and z = -0.83, p = .204 for the continuous test). Furthermore, the observed distribution is flatter than would be expected if the studies were simply underpowered. Statistical analyses confirmed that the *p*-curve was significantly flatter than a null of 33% power (binomial p = .019; continuous test: z = -2.08, p = .019). The results were very similar for the robustness test, which also failed to find significant evidence of right-skewness (p = .500 for the binomial test, and z = -0.88, p = .190 for the continuous test) and, in fact, detected that the *p*-curve was significantly flatter than a 33%-powered *p*-curve, (p = .033 for the binomial test and z = -1.99, p = .023 for the continuous test).

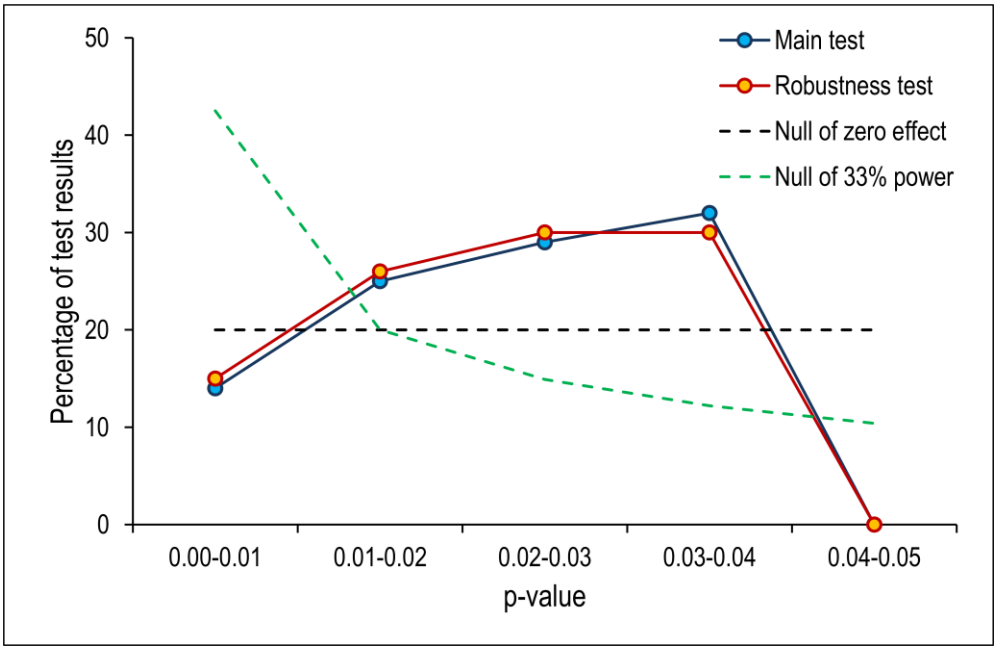

Figure 1. *P*-curve of the key statistical contrasts included in the analysis compared to the expected distribution when the null hypothesis is true (null of zero effect) or when the alternative hypothesis is true but the experiments lack sufficient power (null of 33% power).

It is interesting to note that the results of *p*-curve do not improve if one removes from the analyses all the studies that explored the effects of glucose rinsing, which are based on a somewhat different theoretical background. Neither the continuous test (z = -0.57, p = .285) nor the binomial test (p = .668) suggested that the remaining set of studies had any evidential value after removing glucose rinsing experiments. Furthermore, both tests (continuous z = -1.94, p = .026; binomial p =

.024) suggested that the *p*-curve was significantly flatter than a null of 33% power. In other words, the poor results of the previous tests cannot be attributed to the inclusion of glucose rinsing studies in the analyses.

Discussion

The results of our analyses suggest that the relationship between glucose levels and self-control behaviors might be unreliable. Figure 1 shows that the key *p*-values of the 19 studies included in the present analyses follow a surprisingly flat distribution. This is exactly the pattern of results that one would expect to find if those results were false positives. These results remain unchanged regardless of whether glucose rinsing studies are included or excluded from the sample.

These results may not come as a complete surprise given the challenges the glucose hypothesis has come up against empirically through failed replications (Job et al., 2013; Kelly et al., 2015; Lange & Eggert, 2014; Lange et al., 2014) and theoretical critiques regarding its biological plausibility (Kurzban, 2010; Osman, 2014). Furthermore, a detailed analysis of the seminal paper suggesting the glucose hypothesis showed that the number of significant findings reported in that paper was too large, given the low power of each study (Schimmack, 2012). In other words, the results were "too good to be true" (Francis, 2012). Nevertheless, the findings from the present study are a surprise in the context of the wide acceptance of the glucose hypothesis in general scientific research and its popularity, as evidenced by the number of citations that the original work by Gailliot et al (2007) has received and the continued influence of this hypothesis in recent reviews on ego depletion (e.g., Baumeister, 2014; Baumeister & Alghamdi, in press). Moreover, the hypothesis has intuitive and seemingly practical appeal. If one accepts that the many personal and societal problems we face come down to failing to regulate our actions by not exerting enough self-control (Baumeister et al. 2000), then glucose supplements would provide a simple means to enhance our will power and ameliorate these problems (Baumeister & Tierny, 2011). In light of our results, it is dubious that such a recommendation will work in the real world. This conclusion converges with recent evidence that glucose might have little or no impact on domain general decision-making tasks (Orquin & Kurzban, in press) and an intriguing series of meta-analyses and pre-registered replications suggesting that the ego-depletion effect itself might be less robust than previously thought (Carter, Kofler, Forster, & McCullough, 2015; Hagger et al., in press).

Previous criticisms of the glucose model of ego-depletion have typically focused on individual papers (e.g., Kurzban, 2010; Schimmack, 2012). Article-level analyses like those of Francis (2012) are ideal in some respects because they ensure that all the studies under scrutiny are

grounded in the same theoretical view and rely on very similar research methods. Unfortunately, only a couple of the papers included in the present review contain a sufficiently large number of studies to attempt this type of analysis (Gailliot et al., 2007, and possibly Hagger & Chatzinsarantis, 2013). An examination of the wider literature, like the one offered in the present article, must necessarily collate studies with heterogeneous methods and theoretical backgrounds. In exchange, this approach allows researchers to check for publication and reporting biases in areas of research where articles with a small number of studies are prevalent. In this sense, our study adds to the conclusions of article-level analyses by suggesting that the kind of biases that have been detected in isolated studies might be representative of the wider area of research on the glucose model of ego-depletion. In any case, it is also important to note that, with the possible exception of glucose rinsing studies, the rest of the experiments included in the present analyses share a common theoretical background.

It is worth noting that, as with any other statistical test, *p*-curve is not a flawless indicator of bias (Bishop & Thompson, 2016; Bruns & Ioannidis, 2016; Lakens, 2015). Our results suggest that, on average, these studies have little or no evidential value, but they do not allow us to determine whether the significant results are due to publication bias, selective reporting of outcomes/analysis, and/or *p*-hacking. It is not impossible that some of these studies are exploring small but true effects, and that their evidential value may be diluted by the biases that pervade the rest of the studies. Perhaps future research will show that glucose does play a role in ego-depletion effects, but our conclusions are based on the analysis of the extant literature in this area. Thus, our contribution must be seen as an additional piece of information in the wider context of attempts to verify the reliability of the glucose model of ego-depletion. It is also important to remark that the kind of biases explored in the present study are prevalent in other (but not all) areas of psychological research (e.g., Bakker, van Dijk, & Wicherts, 2012) and that low reproducibility is not exclusively a problem of psychological research (Camerer et al., 2016; Errington, Iorns, Gunn, Tan, Lomax, & Nosek, 2014). In fact, it is fair to say that psychology is taking a leading role in the dissemination of open research practices (Open Science Collaboration, 2015). We hope that this new trend in psychological research will soon render meta-analytic studies like the present one unnecessary.

Author contributions

All authors developed the study concept. The literature search was conducted by M. A. Vadillo and N. Gold. M. A. Vadillo performed the data analysis and interpretation. M. Osman drafted the

manuscript, and M. A. Vadillo and N. Gold provided critical revisions. All authors approved the final version of the manuscript for submission.

Author note

We are indebted to Greg Francis, David R. Shanks, Eric-Jan Wagenmakers, and two anonymous reviewers for their valuable comments on earlier versions of this article. Natalie Gold was supported by the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013; ERC Grant Agreement n. 283849). Correspondence concerning this article should be addressed to Magda Osman, Queen Mary University of London, Biological and Experimental Psychology, School of Biological and Chemical Sciences, Mile End Road, London E1 4NS, United Kingdom. E-mail: m.osman@qmul.ac.uk

References

1. Bakker, M., van Dijk, A., & Wicherts, J. M. (2012). The rules of the game called psychological science. *Perspectives on Psychological Science, 7,* 543–554.

2. Baumeister, R. F. (2014). Self-regulation, ego depletion, and inhibition. *Neuropsychologia, 65,* 313-319.

3. Baumeister, R. F., & Alghamdi, N. G. (in press). Role of self-control failure in immoral and unethical actions. *Current Opinion in Psychology.*

4. 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.

5. Baumeister, R. F., Muraven, M., & Tice, D. M. (2000). Ego depletion: A resource model of volition, self-regulation, and controlled processing. *Social Cognition*, *18*, 130-150.

6. Baumeister, R. F., & Tierny, J. (2011). *Willpower: Rediscovering the greatest human strength.* New York, NY: Penguin Press.

7. * Birnie, L. H. W., Smallwood, J., Reay, J., & Riby, L. M. (2015). Glucose and the wandering mind: Not paying attention or simply out of fuel? *Psychopharmacology*, 232, 2903-2910.

8. Bishop, D. V. M., & Thompson, P. A. (2016). Problems in using p-curve analysis and text-mining to detect rate of p-hacking and evidential value. *PeerJ, 4,* e1715.

9. Bruns, S. B., & Ioannidis, J. P. A. (2016). P-curve and p-hacking in observational research. *PLoS ONE, 11,* e0149144.

10. Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Heikensten, E. (2016). Evaluating replicability of laboratory experiments in economics. *Science,* 351(6277).

11. 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.

12. Carter, E. C., & McCullough, M. E. (2013). After a pair of self-control-intensive tasks, sucrose swishing improves subsequent working memory performance. *BMC Psychology, 1,* 22.

13. * Denson, T. F., von Hippel, W., Kemp, R. I., & Teo, L. S. (2010). Glucose consumption decreases impulsive aggression in response to provocation in aggressive individuals. *Journal of Experimental Social Psychology*, *46*, 1023-1028.

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

15. * Dickinson, D. L., McElroy, T., & Stroh, N. (2014). Impact of glucose on Bayesian versus heuristic-based decision making. *Journal of Neuroscience, Psychology, and Economics*, *7*, 237-247.

16. * Dvorak, R. D., & Simons, J. S. (2009). Moderation of resource depletion in the self-control strength model: Differing effects of two modes of self-control. *Personality and Social Psychology Bulletin, 35,* 572-583.

17. Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., & Nosek, B. A. (2014). An open investigation of the reproducibility of cancer biology research. *eLife, 3,* e04333.

18. Francis, G. (2012). Too good to be true: Publication bias in two prominent studies from experimental psychology. *Psychonomic Bulletin & Review, 19,* 151-156.

19. Freud, S. (1961). The ego and the id. In J. Strachey (Ed.), The standard edition of the complete psychological works of Sigmund Freud (Vol. 19, pp. 12-66). London: Hogarth Press. (Original work published 1923.)

20. * Gailliot, M. T. (2012). Improved self-control associated with using relatively large amounts of glucose: Learning self-control is metabolically expensive. *Psychology*, *3*, 987-990.

21. Gailliot, M. T., & Baumeister, R. F. (2007). The physiology of willpower: Linking blood glucose to self-control. *Personality and Social Psychology Review, 11,* 303-327.

22. * Gailliot, M. T., Baumeister, R. F., DeWall, C. N., Maner, J. K., Plant, E. A., Tice, D. Brewer, L., & Schmeichel, B. J. (2007). Self-control relies on glucose as a limited energy source: willpower is more than a metaphor. *Journal of personality and social psychology*, *92*, 325-336.

23. * Gailliot, M. T., Peruche, B. M., Plant, E. A., & Baumeister, R. F. (2009). Stereotypes and prejudice in the blood: Sucrose drinks reduce prejudice and stereotyping. *Journal of Experimental Social Psychology*, *45*, 288-290.

24. Hagger, M. S., Chatzisarantis, N. L. D., Alberts, H., Anggono, C. O., Batailler, C., Birt, A., … Zwienenberg, M. (in press). A multi-lab pre-registered replication of the ego-depletion effect. *Perspectives on Psychological Science.*

25. * Howard, M. A., & Marczinski, C. A. (2010). Acute effects of a glucose energy drink on behavioral control. *Experimental and Clinical Psychopharmacology*, *18*, 553-561.

26. * Job, V., Walton, G. M., Bernecker, K., & Dweck, C. S. (2013). Beliefs about willpower determine the impact of glucose on self-control. *Proceedings of the National Academy of Sciences*, *110*, 14837-14842.

27. * Kelly, C. L., Sünram-Lea, S. I., & Crawford, T. J. (2015). The role of motivation, glucose and self-control in the antisaccade task. *PloS One*, *10*, e0122218.

28. Kurzban, R. (2010). Does the brain consume additional glucose during self-control tasks? *Evolutionary Psychology, 8,* 244-259.

29. Lakens, D. (2015). What p-hacking really looks like: A comment on Masicampo and LaLande (2012). *Quarterly Journal of Experimental Psychology, 68,* 829-832.

30. * Lange, F., & Eggert, F. (2014). Sweet delusion. Glucose drinks fail to counteract ego depletion. *Appetite*, *75*, 54-63.

31. * Lange, F., Seer, C., Rapior, M., Rose, J., & Eggert, F. (2014). Turn it all you want: Still no effect of sugar consumption on ego depletion. *Journal of European Psychology Students*, *5(3)*, 1-8.

32. * Masicampo, E. J., & Baumeister, R. F. (2008). Toward a physiology of dual-process reasoning and judgment: Lemonade, willpower, and expensive rule-based analysis. *Psychological Science*, *19*, 255-260.

33. * Molden, D. C., Hui, C. M., Scholer, A. A., Meier, B. P., Noreen, E. E., D'Agostino, P. R., & Martin, V. (2012). Motivational versus metabolic effects of carbohydrates on self-control. *Psychological Science*, *23*, 1137-1144.

34. * Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

35. Orquin, J. L., & Kurzban, R. (in press). A meta-analysis of blood glucose effects on human decision making. *Psychological Bulletin.*

36. Osman, M. (2014). *Future-minded: The psychology of agency and control.* Palgrave-MacMillan.

37. Owen, L., Scholey, A. B., Finnegan, Y., Hu, H., & Sünram-Lea, S. I. (2012). The effect of glucose dose and fasting interval on cognitive function: A double-blind, placebo-controlled, six-way crossover study. *Psychopharmacology, 220,* 577-589.

38. * Sanders, M. A., Shirk, S. D., Burgin, C. J., & Martin, L. L. (2012). The gargle effect: Rinsing the mouth with glucose enhances self-control. *Psychological Science*, *23*, 1470-1472.

39. Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*, 551-566.

40. Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 11, 1359-1366.

41. Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, 534-547.

42. * Wang, X. T., & Dvorak, R. D. (2010). Sweet future fluctuating blood glucose levels affect future discounting. *Psychological Science*, *21*, 183-188.

Table 1. List of studies and statistical contrasts included in the analyses

Paper	Study #	Sugar rinsing?	Key statistical contrast	*p*-value
Birnie, Smallwood, Reay, & Riby (2015)		No	$t(15)=2.469$*	.02605
Denson, von Hippel, Kemp, & Teo (2010)	1	No	$t(67)=-2.19$	.03201
	2	No	$t(151)=2.24$	.02655
DeWall, Baumeister, Gailliot, & Maner (2008)	2	No	$F(1, 55) = 6.64$	.01268
Dickinson, McElroy, & Stroh (2014)		No	$z = 1.88$	.06011
Dvorak & Simons (2009)		No	$F(1, 177) = 5.63$	.01873
Gailliot (2012)		No	$r(50) = -.30$	.03071
Gailliot, Baumeister, DeWall, Maner, Plant, Tice, Brewer, & Schmeichel (2007)	1	No	$F(1, 100) = 6.08$	.01537
	2	No	$t(33) = 2.20$	.03492
	3	No	$r(14) = -.62$*	.01041
	4	No	$r(10) = .56$	.05828
	5	No	$r(21) = .45$	.03120
	6	No	$r(15) = .43$	.08493
	7	No	$F(1, 57) = 5.04$	.02866
	8	No	$F(1, 69) = 5.45$	.02249
	9	No	$t(16) = 3.13$	.00646
Gailliot, Peruche, Plant, & Baumeister (2009)		No	$t(47) = 2.21$*	.03201
Hagger & Chatzinsarantis (2013)	1	Yes	$F(1, 24) = 8.42$	.00783
	2	Yes	$F(1, 30) = 6.12$	.01925
	3	Yes	$F(1, 32) = 4.06$	.05238
	4	Yes	$F(1, 40) = 10.32$	.00260
	5	Yes	$F(1, 36) = 7.28$	.01055
Howard & Marczinski (2010)		No	$F(4, 75) = 2.95$	.02544
Job, Walton, Bernecker, & Dweck (2013)	1	No	$t(78) = 2.10$	.03896
	2	No	$F(1,58) = 5.16$	.02684
	3	No	$F(1,139) = 5.28$	.02306

Kelly, Sünram-Lea, & Crawford (2015)		No	$F(1, 67) = 0.80$	.37430
Lange & Eggert (2014)	1	No	$F(1,68) = 1.12$	.29366
	2	No	$F(1,110) = 0.01$*	.92053
Lange, Seer, Rapior, Rose, & Eggert (2014)		No	$t(68) = 0.05$*	.96027
Masicampo & Baumeister (2008)		No	$F(1, 111) = 5.311$	.02305
Molden, Hui, Scholer, Meier, Noreen, D'Agostino, & Martin (2012)	1	No	$F(1, 83) = 2.05$	.15596
	2	Yes	$F(1, 39) = 4.54$	.03947
	3	Yes	$F(1, 28) = 5.02$	.03317
Open Science Collaboration (2015). Replication of Masicampo & Baumeister (2008)		No	$F(1,158) = 0.379$	.53902
Sanders, Shirk, Burgin, & Martin (2012)		Yes	$t(49) = -2.129$	.03831
Wang & Dvorak (2010)		No	$t(31) = 2.55$	.01593
		No	$t(32) = 3.12$	.00381

Note. The statistical contrasts were selected following the guidelines of Simonsohn et al. (2014). A complete *p*-curve disclosure table justifying the selection of each statistical contrast is available at the Open Science Framework http://bit.ly/1PoHFby. The statistical contrasts marked with asterisks were replaced by alternative contrasts in the robustness test (see the online *p*-curve disclosure table for further details).