

Many Labs 5: Registered Replication of Vohs and Schooler (2008), Experiment 1



**Nicholas R. Buttrick¹, Balazs Aczel², Lena F. Aeschbach³,
Bence E. Bakos², Florian Brühlmann³, Heather M. Claypool⁴,
Joachim Hüffmeier⁵, Marton Kovacs², Kurt Schuepfer⁴,
Peter Szecsi², Attila Szuts², Orsolya Szöke², Manuela Thomae^{6,7},
Ann-Kathrin Torka⁵, Ryan J. Walker⁴, and Michael J. Wood⁸**

¹Department of Psychology, University of Virginia; ²Institute of Psychology, ELTE Eötvös Loránd

University; ³Department of Psychology, University of Basel; ⁴Department of Psychology, Miami

University; ⁵Department of Psychology, TU Dortmund University; ⁶MEU - Die Multiversität;

⁷Diploma University of Applied Sciences; and ⁸Department of Psychology, University of Winchester

Abstract

Does convincing people that free will is an illusion reduce their sense of personal responsibility? Vohs and Schooler (2008) found that participants reading from a passage “debunking” free will cheated more on experimental tasks than did those reading from a control passage, an effect mediated by decreased belief in free will. However, this finding was not replicated by Embley, Johnson, and Giner-Sorolla (2015), who found that reading arguments against free will had no effect on cheating in their sample. The present study investigated whether hard-to-understand arguments against free will and a low-reliability measure of free-will beliefs account for Embley et al.’s failure to replicate Vohs and Schooler’s results. Participants ($N = 621$) were randomly assigned to participate in either a close replication of Vohs and Schooler’s Experiment 1 based on the materials of Embley et al. or a revised protocol, which used an easier-to-understand free-will-belief manipulation and an improved instrument to measure free will. We found that the revisions did not matter. Although the revised measure of belief in free will had better reliability than the original measure, an analysis of the data from the two protocols combined indicated that free-will beliefs were unchanged by the manipulations, $d = 0.064$, 95% confidence interval = $[-0.087, 0.22]$, and in the focal test, there were no differences in cheating behavior between conditions, $d = 0.076$, 95% CI = $[-0.082, 0.22]$. We found that expressed free-will beliefs did not mediate the link between the free-will-belief manipulation and cheating, and in exploratory follow-up analyses, we found that participants expressing lower beliefs in free will were not more likely to cheat in our task.

Keywords

attitudes, beliefs, decision making, goals, morality, multilevel modeling, R, direct replication, open data, open materials, preregistered

Received 4/13/18; Revision accepted 3/16/20

What happens when people believe that, for whatever reason, their actions are not wholly their own—that their will is not completely free? Do they use that unshackling as an excuse to behave immorally, or are the moral consequences less dire? Given current findings that individual behavior is influenced by such factors as one’s genome (e.g., Caspi et al., 2002), one’s epigenome (e.g., Bird, 2007), one’s socioecological environment (e.g., Oishi, 2014), and the structural

factors of the society in which one lives (e.g., Gelfand et al., 2011), how do people interpret the concept that total free will, in which any choice can be made at any

Corresponding Author:

Nicholas R. Buttrick, Department of Psychology, University of Virginia,
Gilmer Hall, P.O. Box 400400, Charlottesville, VA 22904-4400
E-mail: nrb8pv@virginia.edu

time, is, in some sense, illusory—that we are all bound by nature, culture, and environment?

Free will, or at least the ability to act other than one has acted, is generally thought to be a precondition for holding people accountable for their moral or immoral actions, both by moral philosophers (e.g., Aristotle, trans. 1980; Kant, 1785/1998) and by ordinary laypeople (e.g., Nichols & Knobe, 2007). If one cannot act freely, then the blameworthiness of one's actions becomes less clear. If free will appears to be constrained, as suggested by arguments that psychological states are biologically rooted, psychopathic offenders, for example, can seem less culpable, even to highly trained U.S. state trial judges (Aspinwall, Brown, & Tabery, 2012).

Given the strong intuitions people have about the relationship between free will and moral action, how do people interpret the moral status of their own activity if they learn that their own actions are not fully free? If people understand contemporary science as indicating that individuals do not have free will, and consequently come to believe that they are no longer fully responsible for their actions and thus can act with less fear of sanction (a set of beliefs that match the existentialist concept of *bad faith*, e.g., Bakewell, 2017), then scientists may be the vectors for a wave of immoral behavior.

Vohs and Schooler (2008) examined this troubling possibility. They predicted that participants led to believe that scientists had concluded that free will is an illusion would find it easier to cheat on experimental tasks than would participants in a control condition. In their first of two studies, they had participants read one of two essays drawn from the same source. In one essay, the author wrote about advances in neuroscience, and in the other, the author discussed findings indicating that free will is an illusion. All participants then completed a math task rigged to allow them to cheat by illicitly looking at the correct answer. Participants given the anti-free-will passage cheated more often than did those given the neuroscience passage. This pattern was mediated by scores on the measure of belief in free will.

This study was the focus of one replication (Embley, Johnson, & Giner-Sorolla, 2015) in the Reproducibility Project: Psychology (RP:P; Open Science Collaboration, 2015). Embley et al. (2015) conducted a replication of Vohs and Schooler's (2008) Experiment 1, using the same experimental design, but found no differences in cheating behavior between the two experimental conditions, $d = 0.20$, 95% confidence interval (CI) = $[-0.33, 0.74]$, $p = .44$, and no relationship between free-will beliefs and cheating, $r = -.05$, 95% CI = $[-.31, .22]$, $p = .70$. In this replication, however, the free-will manipulation also failed to change measured beliefs in free

will, $d = -0.29$, 95% CI = $[-0.83, 0.24]$, $p = .28$. This failure, along with the markedly low reliability of the free-will-belief measure ($\alpha = .43$), means that one should be cautious in interpreting the findings.

It is possible that Embley et al. (2015) were not able to replicate the effect Vohs and Schooler (2008) observed simply because they failed to change participants' beliefs about free will in the first place. If participants in the anti-free-will condition did not come to believe that free will is an illusion, one would not expect their cheating to be affected by the manipulation. In the absence of a successful manipulation, or a reliable measure of the free-will construct, the replication study does not shed much light on the relationship between free-will beliefs and moral behavior.

In this project, we compared results obtained replicating Vohs and Schooler's (2008) Experiment 1 using the methods and materials of Embley et al. (2015) and using a revised protocol. Participants were randomly assigned either to Embley et al.'s RP:P protocol or to our revision, in a 2 (protocol: RP:P vs. revised) \times 2 (condition: control vs. anti-free-will) fully crossed design.

The primary focus for the revision was to re-create the psychological states of participants in the original study. Therefore, in our revised protocol, we replaced the anti-free-will manipulation used by Embley et al. (2015) and by Vohs and Schooler (2008) with a version that was designed to be easier to understand, and we replaced the measure of free-will beliefs used in those studies with an updated version of the scale, designed to have better reliability. By doing so, we hoped to better test the claim that decreasing belief in free will leads to increased cheating.

Disclosures

Preregistration

Our design and confirmatory analyses were preregistered on the Open Science Framework at <https://osf.io/peuch/>.

Data, materials, and online resources

All materials, translations, data files, and analysis scripts are available on the Open Science Framework at <https://osf.io/8rcbk/>. All differences between the original Stage 1 manuscript and the final accepted version are reported at <https://osf.io/pje7s/>. The text of our free-will-belief manipulation, information about the pretest of this manipulation, and supplementary results (e.g., exploratory models excluding suspicious participants, meta-analytic estimates of the effect that include all previous instantiations of this paradigm) can be found in the

Table 1. Demographics of the Sample, by Site and Protocol

Site	Country	Language	Many Labs 5 Protocol			RP:P protocol		
			<i>n</i>	Mean age	Female (%)	<i>n</i>	Mean age	Female (%)
University of Basel	Switzerland	German	71	23.32 (5.07)	81.69	78	23.01 (4.81)	78.20
TU Dortmund University	Germany	German	70	21.29 (2.92)	92.86	69	22.46 (5.68)	89.86
ELTE Eötvös Loránd University	Hungary	Hungarian	57	—	—	58	—	—
Miami University	United States	English	72	19.01 (1.22)	58.33	74	18.86 (1.07)	66.22
University of Winchester	United Kingdom	English	72	21.21 (4.26)	90.28	—	—	—

Note: Values in parentheses are standard deviations. Information on age and gender was not recorded at ELTE Eötvös Loránd University. At the University of Winchester, the Reproducibility Project: Psychology (RP:P) protocol was not used.

Supplemental Material (<http://journals.sagepub.com/doi/suppl/10.1177/2515245920917931>).

Reporting

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Ethical approval

Data were collected in accordance with the Declaration of Helsinki and with approval from the University of Virginia's institutional review board (SBS# 2016-0294).

Method

Participants

We recruited participants at five sites, in five separate countries. In the original study, the value of *d* for the effect of interest was 0.88.¹ On the basis of this effect size, we calculated that 70 participants per site per protocol were required to achieve 95% power. Four sites collected data from participants using both protocols, and the fifth, because of the size of the participant pool, collected data using only the revised protocol. The final sample consisted of 621 individuals. Table 1 summarizes the sample's demographics.

Materials and procedure

Overview. Participants completed one of two protocols designed to test whether decreased belief in free will lead to increased cheating. At sites able to collect data for both protocols, participants were randomly assigned both to protocol and to condition, in a fully crossed 2 (protocol: RP:P vs. revised) × 2 (condition: control vs. anti-free will) design. At the one site able to collect data using only one protocol (the revised protocol), participants were randomly assigned to condition only. The two protocols had

the same basic outline: Participants came to the lab individually and were randomly assigned to either an anti-free-will or a control condition.² Those in the anti-free-will condition were given reading material challenging their belief in free will, whereas those in the control condition were given neutral reading material. After reading the assigned passage, participants were asked to fill out a measure of their free-will beliefs, which asked about individuals' control over and culpability for their actions, and a measure of their positive and negative mood.

Subsequently, participants were given a computerized math task, in which they were supposed to solve 20 multistage arithmetic problems without the help of a calculator.³ They were told that there was a glitch in the program and that

the answer pops up accidentally, unless you hit the SPACEBAR as soon as you see the problem. But we can't tell if the answer ever comes up or not, so we need your help: If you hit the spacebar, then the problem remains on the screen, without the answer, and you can take your time in answering it. But because we don't know whether the answer appeared or not, it is important that you hit the spacebar right away so that the experiment is conducted properly.

In actuality, the program had been rigged not only to show the answers, but also to record the number of space-bar presses. Following Vohs and Schooler (2008), we operationalized cheating as the number of problems for which participants failed to press the space bar to prevent the answer from appearing.

After completing this task, all participants were asked to answer an open-ended question on a separate sheet of paper. The question asked if they thought there was anything unusual about the study and what they thought the study was about. Responses were coded for suspicion by two researchers blind to condition, protocol, and participants' math-task performance.⁴ Participants were then debriefed.

At sites with mainly non-English-speaking populations, all materials were translated into the local language by one member of the research team and then back-translated by a different member. The translations were inspected by the lead author, and all discrepancies were collectively resolved. All materials, including the translations, are available at <https://osf.io/3umks/>.

Difference between the original study and both protocols. In the review process, we discovered that the affect measure used by us and by Embley et al. (2015) was not in fact the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988), which was originally used by Vohs and Schooler (2008), but was instead an ad hoc affect measure of uncertain provenance. Unlike the PANAS, this measure asked about participants' happiness, sadness, euphoria, disgust, pleasure, joy, anger, fear, admiration, love, remorse, guilt, hope, shame, resentment, and tenderness. Therefore, we consider the protocol based on the materials of Embley et al. to be a close replication, rather than a direct one.

Differences between the protocols. The major difference between the protocols was in the details of the manipulation of free-will beliefs and in the scale used to measure these beliefs. Those changes are described in the paragraphs that follow. Additionally, the manipulation and free-will-beliefs measure were presented on paper in the RP:P protocol and on a computer in the revised protocol.

Anti-free-will-belief induction. In the original study and in the RP:P protocol, the free-will beliefs of participants were manipulated by asking them to carefully read a page-long passage drawn from *The Astonishing Hypothesis* (Crick, 1994). In the anti-free-will condition, the passage was taken from a postscript to the book and argued that free will is an illusion. In the control condition, the passage was taken from the middle of the book and discussed consciousness, without any free-will content. It is entirely plausible that Vohs and Schooler's (2008) finding was not replicated by Embley et al. (2015) because participants at the replication site were unable or unwilling to sufficiently engage with the difficult material. To make it easier for participants to understand the assigned passage, in the revised protocol we used a variant of Vohs and Schooler's induction developed by Alquist, Ainsworth, and Baumeister (2013). As their belief manipulation, Alquist et al. had participants read and then paraphrase 10 sentences. Among their conditions were one in which the 10 sentences were all against the idea of free will and another in which the sentences had no free-will content. Every sentence in the anti-free-will condition was drawn from the same Crick (1994) passage

used by Vohs and Schooler (e.g., "Science has demonstrated that free will is an illusion," and "Everything a person does is a direct consequence of their environment and genetic makeup"); the control sentences were drawn from encyclopedia articles (e.g., "Sugar cane and sugar beets are grown in 112 countries," and "Alkaline power cells generally work longer than ordinary batteries"). The full text of the induction materials is provided in the Supplemental Material.

Measure of free-will beliefs. In the original study (Vohs & Schooler, 2008) and in Embley et al.'s (2015) replication, free-will beliefs were measured using the seven-item Free Will subscale of the Free Will and Determinism (FWD) scale (Paulhus & Margesson, 1994). Embley et al. found that this scale had unacceptably low reliability ($\alpha = .43$) that rendered it an unfit measure of free-will beliefs. Luckily, the lead author of the original FWD scale has since developed an updated instrument to measure the same construct: the FAD-Plus (Paulhus & Carey, 2011a). The new eight-item subscale of the FAD-Plus that measures belief in free will has better reported reliability than the version used by Vohs and Schooler ($\alpha = .70$). Because the new measure has greater reliability and taps into the same construct (Paulhus & Carey, 2011b), we used it in the revised protocol.

Pretesting the revised induction and measurement scale. A 200-person pretest, prior to the initiation of the present study, found that the revised belief-in-free-will scale had acceptable reliability ($\alpha = .84$), and that the revised anti-free-will-belief induction seemed to successfully decrease free-will beliefs relative to the revised control condition, as measured by the revised belief-in-free-will scale, $d = 0.36$, 95% CI = [0.07, 0.64], $p = .013$ (see the Supplemental Material for details).

Results

Following the original study, we had no data-exclusion rule.

Overall, we did not find good reliability for the free-will measure, $\alpha = .45$, 95% CI = [.39, .52], for the RP:P protocol, and $\alpha = .72$, 95% CI = [.69, .76], for the revised protocol. Alphas for this measure did not differ by site for either the RP:P protocol, $\chi^2(3, N = 4) = 3.67$, $p = .30$, or the revised protocol, $\chi^2(4, N = 5) = 2.69$, $p = .61$ (Feldt, Woodruff, & Salih, 1987). Because the affect measure was equivalent across protocols, we collapsed the data across protocols when we calculated the reliability of this measure. We found that it had good reliability for both positive affect, $\alpha = .85$, 95% CI = [.83, .87], and negative affect, $\alpha = .82$, 95% CI = [.80, .84]. Alphas for positive affect did not differ by site, $\chi^2(4,$

$N = 5$) = 3.33, $p = .51$. In contrast, alphas for negative affect did differ by site, $\chi^2(4, N = 5) = 9.47$, $p = .050$; alphas were lower for TU Dortmund University ($\alpha = .75$, 95% CI = [.68, .81]) and the University of Basel ($\alpha = .76$, 95% CI = [.69, .81]) relative to the other sites (see the Supplemental Material for by-site plots of all alphas).

Structure of the analyses

Because our data have participants nested within protocol and site, we fitted multilevel models to test our hypotheses (e.g., Raudenbush & Bryk, 2002). Multilevel models with complicated data structures often do not converge with relatively small samples (e.g., Bell et al., 2010), so we set up a hierarchy of random-effects terms in case our models were statistically unidentifiable. In all models, we started out allowing the interaction of protocol and condition to vary across sites (setting a random slope for the protocol-by-condition interaction and a random intercept for site). If that model did not converge, we dropped the random slope for protocol, freeing just the effect of condition to differ across sites (setting a random slope for condition and a random intercept for site). If that model did not converge, we simply allowed sites to have their own random intercepts. In analyses with tests run on the two protocols in parallel, we used the most complex random-effects term that converged in both models. Analyses were run using the *lme4* (Bates, Mächler, Bolker, & Walker, 2015), *mediation* (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014), and *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017) packages in R, and all p values for the multilevel tests are Satterthwaite approximations. Our analysis scripts can be found at <https://osf.io/wzu39/>.

Confirmatory analyses

Manipulation check. To test whether the manipulation of free-will beliefs was successful, we first standardized the free-will scores within protocol, to create a comparable metric across the data sets. We predicted these free-will scores with the interaction of protocol and condition as a fixed effect, and with a random slope for condition (across sites) and a random intercept for site. We found that protocol did not interact with condition, $b = 0.28$, $SE = 0.16$, $t(608.96) = 1.77$, $p = .077$. Because we found no interaction with protocol, we collapsed the data across protocols, predicting free-will scores from condition and a random slope for condition (across sites) and a random intercept for site. We found that condition did not affect free-will beliefs (control condition: $M = -0.04$, $SD = 0.98$; anti-free-will condition: $M = 0.04$, $SD = 1.02$), $b = 0.065$, $SE = 0.085$, $t(14.01) = 0.77$, $p = .46$, $d = 0.064$, 95% CI = [-0.087, 0.22].⁵

Affect effects. We did not expect the manipulation to alter participants' affect. To test our prediction, we fitted two models, one predicting positive affect and one predicting negative affect. The models included the interaction of protocol and condition as a fixed effect, a random slope for condition (across sites), and a random intercept for site. Condition did not interact with protocol for either positive affect, $b = 0.17$, $SE = 1.02$, $t(579.28) = 0.17$, $p = .87$, or negative affect, $b = -0.93$, $SE = 0.62$, $t(607.09) = -1.50$, $p = .13$, so for both models, we collapsed the data across protocols. We found that positive affect did marginally differ between conditions; participants in the control condition felt happier ($M = 20.23$, $SD = 6.47$) than did those in the anti-free-will condition ($M = 18.84$, $SD = 6.18$), $b = -1.39$, $SE = 0.59$, $t(7.02) = -2.34$, $p = .052$, $d = -0.22$, 95% CI = [-0.37, -0.057]. Negative affect did not differ between conditions (control condition: $M = 11.10$, $SD = 3.95$; anti-free-will condition: $M = 11.05$, $SD = 4.02$), $b = -0.0076$, $SE = 0.34$, $t(11.79) = -0.022$, $p = .98$, $d = -0.0019$, 95% CI = [-0.16, 0.17].

Focal test. Following Vohs and Schooler (2008), we hypothesized that decreased belief in free will would lead to more cheating on the math task. We fitted a model predicting the number of problems on which a participant cheated from the fixed interaction of protocol and condition, with a random slope for condition (across sites) and a random intercept for site. We found that protocol did not interact with condition, $b = -0.68$, $SE = 1.02$, $t(455.06) = -0.67$, $p = .51$. Figure 1 shows the effect of condition within each protocol at each site.

Because we found no interaction with protocol type, we collapsed the data across protocols, predicting cheating scores from condition, a random slope for condition (across sites), and a random intercept for site. We found that condition did not affect cheating: On average, participants in the control condition cheated on 7.65 problems ($SD = 6.45$), and participants in the anti-free-will condition cheated on 8.18 problems ($SD = 6.59$), $b = 0.50$, $SE = 0.67$, $t(4.37) = 0.744$, $p = .50$, $d = 0.076$, 95% CI = [-0.082, 0.22].

Mediation by free-will beliefs. We expected free-will beliefs to mediate the effect of condition on cheating. Because the *mediation* package does not allow for direct tests of moderated mediation with random-effects models, we built a separate model for each protocol. Allowing for just a random intercept for site (because of package constraints), we did not find evidence for mediation (ab path) in either the RP:P protocol, mediation effect = -0.011, 95% CI = [-0.25, 0.20], $p = .94$, or the revised protocol, mediation effect = 0.0098, 95% CI = [-0.073, 0.13], $p = .86$. Accounting for free-will beliefs did not change the effect of condition on cheating in the

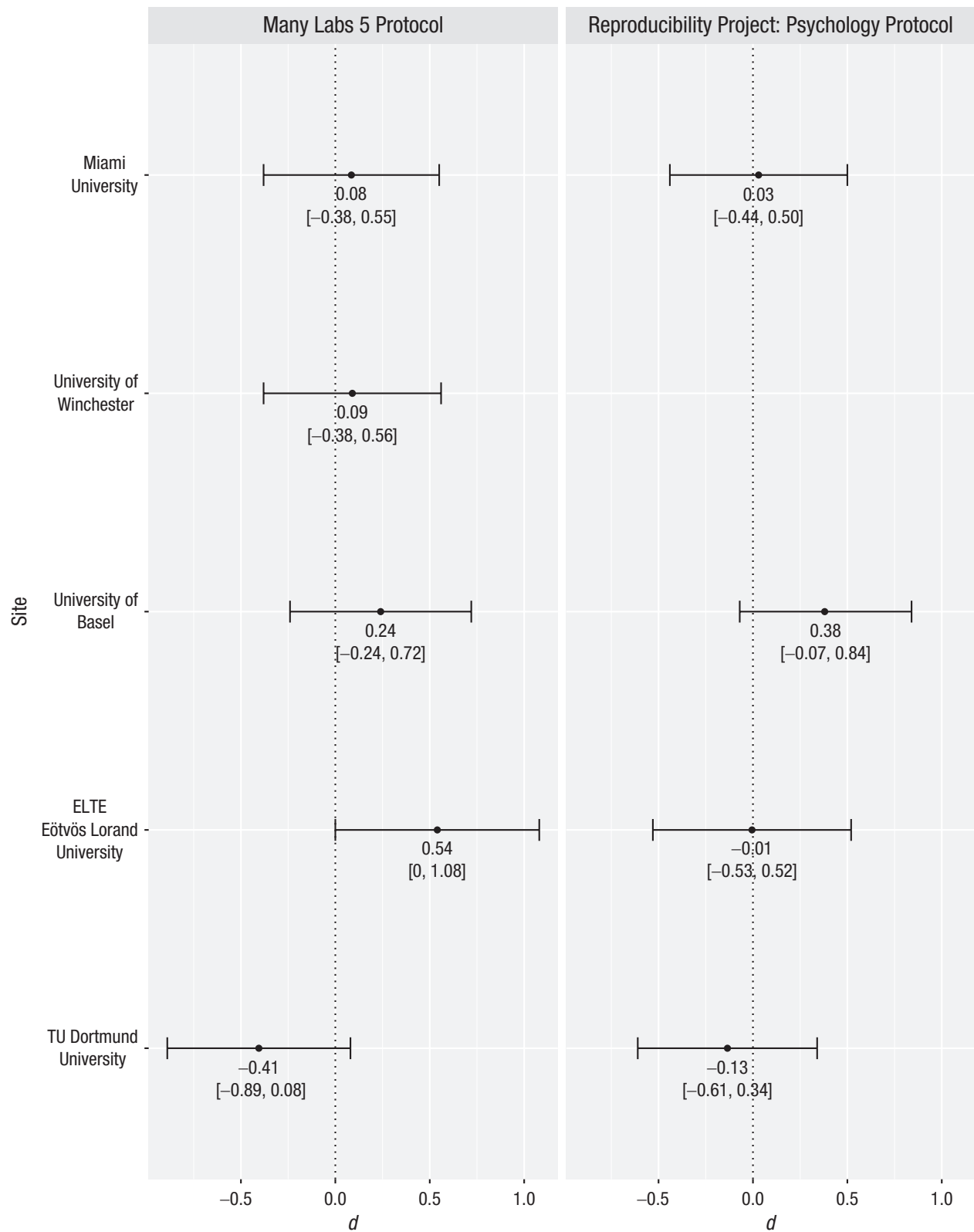


Fig. 1. Effect of condition at each site within each protocol. Point estimates are Cohen's d values, and whiskers indicate 95% confidence intervals (also presented in brackets). Positive values indicate more cheating in the anti-free-will condition than in the control condition.

RP:P protocol, total effect (c path) = 0.12, 95% CI = [-1.40, 1.68], $p = .87$; direct effect (c' path) = 0.13, 95% CI = [-1.45, 1.74], $p = .87$. Accounting for free-will beliefs also did not change the effect of condition on cheating in the revised protocol, total effect (c path) = 0.80, 95% CI = [-0.59, 2.02], $p = .23$; direct effect (c' path) = 0.79, 95% CI = [-0.59, 2.01], $p = .23$. A mediation model collapsing the data across the two protocols similarly did not yield evidence for mediation of the effect of condition on cheating through beliefs about free will: mediation effect (ab path) = -0.013, 95% CI = [-0.092, 0.04], $p = .67$; total effect (c path) = 0.45, 95% CI = [-0.50, 1.36], $p = .35$; direct effect (c' path) = 0.46, 95% CI = [-0.49, 1.36], $p = .34$.

A mediation analysis matching the strategy of the original study led to a similar conclusion (see the Supplemental Material for details).

Exploratory analyses

Suspicion of the dependent variable. In an exploratory analysis, we investigated whether participants' suspicion of the purpose of the math task (i.e., that the program was not actually miscoded but was intended to show them the answers) moderated the focal effect or the mediation. At each site, two independent coders, blind to condition, read participants' comments and assessed whether the participants believed that their behavior was under observation (they did not believe the "glitch" was real or thought it was "part of the experiment"; they thought the study was about honesty, integrity, morality, or cheating; or they thought it was about whether people would really do the math problems or just wait for the answers). We did not code as suspicious anyone who generally thought that the study was strange (they thought it was weird, odd, or unusual that the answers were provided or that they were told not to look at the answers; they were not sure if the provided answers were correct). Between-coder agreement was high, $\kappa = .904$ (Cohen, 1960; Grant, Button, & Snook, 2017), and all disagreements were resolved by the team leads at each site.

Overall, 23.19% of participants reported some suspicion of the manipulation (23.81% of the participants in the control condition, 22.55% of the participants in the anti-free-will condition). To test whether suspicion moderated the effect of free-will beliefs on cheating, we fitted a model predicting cheating from the fixed-effect interaction of condition, protocol, and suspicion, with a random slope for condition (across sites) and a random intercept for site. In this model, we found a significant two-way interaction between condition and suspicion, $b = -3.34$, $SE = 1.60$, $t(554.66) = -2.09$, $p = .037$. Decomposing the interaction via the analysis of marginal means, we found that participants who were

not suspicious of the manipulation were directionally more likely to cheat in the anti-free-will condition ($M = 8.58$, $SD = 6.76$) than in the control condition ($M = 7.62$, $SD = 6.47$), $b = -1.05$, $SE = 0.68$, $t(5.29) = -1.55$, $p = .18$; in contrast, among participants who were suspicious of the manipulation, those in the anti-free-will condition were directionally less likely to cheat ($M = 6.80$, $SD = 5.82$) than were those in the control condition ($M = 7.73$, $SD = 6.44$), $b = 1.45$, $SE = 1.15$, $t(34.75) = 1.26$, $p = .22$ (see Fig. 2).

Given this interaction, we reran all analyses with just those participants who reported no suspicion of the manipulation. Our conclusions largely did not change, but we did find a marginal effect of condition on cheating behavior in the predicted direction, $d = 0.17$, 95%

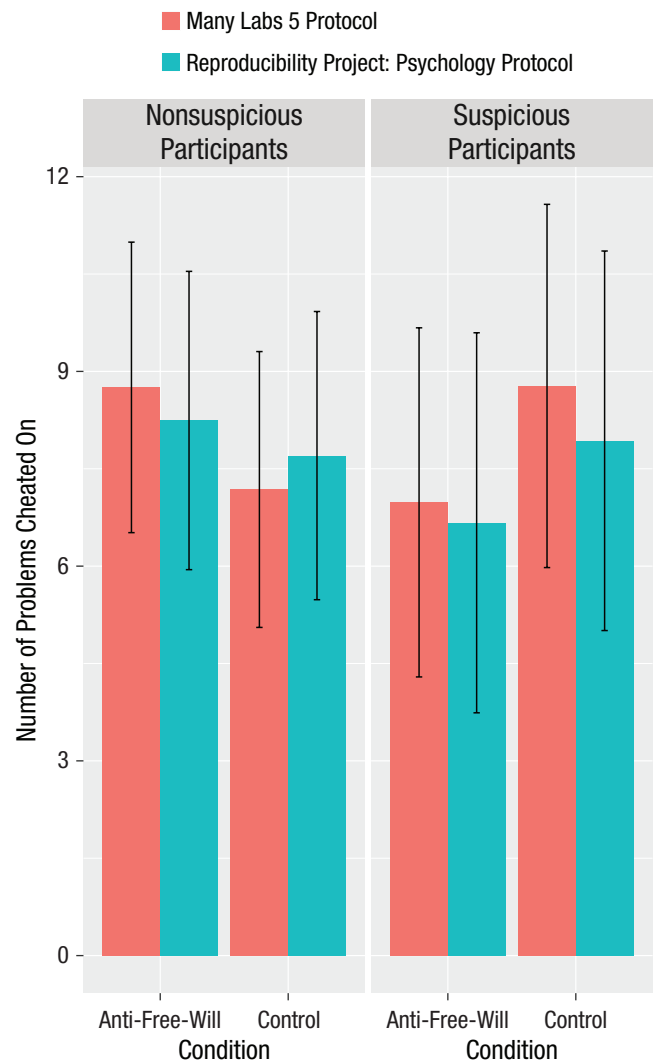


Fig. 2. Cheating scores of participants who were suspicious of the manipulation (right panel) and those who were not suspicious (left panel), by condition and protocol. Error bars are 95% confidence intervals, and estimates come from marginal fixed effects.

CI = [0.004, 0.34], $p = .067$ (see the Supplemental Material for details).

Free-will beliefs and cheating. Because the manipulation did not seem to affect free-will beliefs, we ran an additional set of models to see if free-will beliefs, collapsed across condition, predicted cheating by themselves. We ran a model in which cheating was predicted by the interaction of free-will beliefs with protocol, with a random slope for protocol (across sites) and a random intercept for site. We found no interaction between free-will beliefs and protocol, $b = 0.16$, $SE = 0.51$, $t(610.81) = 0.31$, $p = .75$, so we collapsed the data across protocols. We found no evidence for a relationship between free-will beliefs and cheating, $b = -0.16$, $SE = 0.26$, $t(615.49) = -0.62$, $p = .54$.

Discussion

Does decreasing free-will beliefs lead to an increase in cheating behavior? In their original study, Vohs and Schooler (2008) presented evidence that reading a short argument that free will is an illusion decreased participants' belief about the existence of free will and led to an immediate increase in cheating behavior when participants were subsequently given a (rigged) set of math problems to solve. In an initial replication attempt, Embley et al. (2015) presented the same experimental materials to a new set of participants, but were unable to find evidence for a link between reading the argument and cheating. Participants in that replication, however, may not have been convinced that free will is an illusion, as beliefs about free will did not differ between the experimental and control condition. Thus, it is hard to interpret the null results.

In an attempt to resolve this issue, we designed a revised test of the hypothesis. We simplified the manipulation intended to convince participants of the illusory nature of free will (using materials from Alquist et al., 2013), and we used an improved measure of free-will beliefs (Paulhus & Carey, 2011a). After preregistering the design and analyses, we recruited 621 participants from five sites in five different countries and randomly assigned them to participate in either the original version of the study or our newly revised variant. At one site, where the participant pool was more limited, participants participated in the newly revised variant only.

We did not find evidence supporting the hypothesis. We found no effect of the free-will manipulation on cheating behavior, $d = 0.076$, 95% CI = [-0.082, 0.22], and this finding did not differ between the original protocol and the revised protocol. A random-effects meta-analysis combining this study and those of Vohs and Schooler (2008) and Embley et al. (2015) suggested

that the overall effect size across these studies was not significantly different from zero, $d = 0.14$, 95% CI = [-0.036, 0.31], $p = .12$ (see the Supplemental Material for details). Our results were fairly consistent across our sites: The heterogeneity between sites explained only a small proportion of the variance, adjusted intra-class correlation coefficient = .11 (Johnson, 2014), and the effect of condition was significant for only one protocol at one site (the revised protocol, at ELTE Eötvös Loránd University, $p = .049$).

As did Embley et al. (2015), we found that participants' minds were unchanged by our anti-free-will manipulation, and, in exploratory analyses, we did not find that participants who expressed less belief in free will, measured with either of our free-will belief scales, were any more likely to cheat on our math task. Although the finding that participants with lower free-will beliefs were not more likely to act unethically raises questions about the basic hypothesis of interest, our inability to experimentally manipulate the psychological state of interest renders this an unfit test.

The operationalization of cheating was the one element unchanged across all tests of the hypothesis, and it may be that participants simply did not believe our cover story or, having had prior experience with psychology studies, simply were wary of behaving badly in the lab. They may have suspected that our primary dependent variable was not a measure of their math ability, and may have even suspected that we were interested in measuring their moral behavior. If our dependent variable did not properly operationalize the construct of immoral behavior, then our test's informativeness is quite limited.

Overall, we did see that participants were at least somewhat likely to act unethically in our paradigm, cheating on an average of 7.91 ($SD = 6.52$) problems out of 20 (only 8.7% of participants never cheated at all). This is roughly equivalent to the level of cheating observed by Embley et al. (2015), who found that their participants cheated on an average of 6.05 problems. In their study and in ours, the level of cheating was lower than observed by Vohs and Schooler (2008), who found that participants cheated on an average of 11.84 problems. In an exploratory analysis, however, we did find that participants who expressed suspicion of the manipulation behaved significantly differently from those who did not; naive participants were marginally more likely to cheat when given materials suggesting that free will is an illusion than when given neutral materials. The observed effect size for the test of the condition effect within the subsample of naive participants was fairly small, $d = 0.17$, 95% CI = [0.004, 0.34], roughly the same magnitude as the overall difference between conditions found by Embley et al. ($d = 0.20$),

and far smaller than the effect size reported by Vohs and Schooler ($d = 0.88$).

Using a different set of independent and dependent measures could possibly provide a stronger test of the relationship between believing in free will and ethical behavior. A large correlational study, using data from 46 countries in the World Values Survey (more than 65,000 individuals) found a relationship between disbelief in free will and tolerance of unethical behavior (Martin, Rigoni, & Vohs, 2017). Moreover, recent experimental studies, using different operationalizations of ethics, suggest, for example, that reducing belief in free will by exposing participants to information about neuroscience increases their selfishness in an economic game (Protzko, Ouimette, & Schooler, 2016) and reduces their support for the punishment of a rehabilitated violent criminal (Shariff et al., 2014). In future tests of this or related hypotheses, researchers may want to use one of these alternate approaches to inducing disbelief in free will and measuring immorality, instead of the methods used in the present study.

Finally, in the review process, we discovered that the mood measure used in both Embley et al.'s (2015) replication and in both of our replication protocols differed from that used in Vohs and Schooler's (2008) original study. As the mood measure was presented to participants between the independent and dependent variables, it is possible that this departure from the original protocol is in part responsible for the difference in findings between Vohs and Schooler's study and the three ensuing replication attempts.

Conclusion

Although one individual study is only moderately informative about any given psychological phenomenon (Open Science Collaboration, 2015), there appears to be limited evidence that the manipulations of free-will beliefs described in this article lead to any changes in unethical behavior as operationalized in the present study. Although our exploratory analyses revealed no correlational evidence for the relationship between expressed free-will beliefs and cheating on our task, we can say very little about the causal relationship between belief in free will and willingness to behave unethically, given that we were unable to induce the required psychological state.

Transparency

Action Editor: Daniel J. Simons

Editor: Daniel J. Simons

Author Contributions

N. R. Buttrick designed the study, conducted the analyses, and drafted the manuscript. F. Brühlmann assisted with

the analyses. All the authors except N. R. Buttrick collected the data and revised the manuscript.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

Open Data: <https://osf.io/kuagw/>

Open Materials: <https://osf.io/3umks/>


Preregistration: <https://osf.io/peuch/>


All data and materials have been made publicly available via the Open Science Framework and can be accessed at <https://osf.io/kuagw/> and <https://osf.io/3umks/>, respectively. The design and analysis plans were preregistered at the Open Science Framework and can be accessed at <https://osf.io/peuch/>. The complete Open Practices Disclosure for this article can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2515245920917931>. This article has received badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Nicholas Buttrick  <https://orcid.org/0000-0002-1165-8938>

Florian Brühlmann  <https://orcid.org/0000-0001-8945-3273>

Attila Szuts  <https://orcid.org/0000-0002-7993-1988>

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/2515245920917931>

Notes

1. This effect size was calculated using the means and standard deviations reported in Vohs and Schooler (2008). A calculation based on the reported t value, however, yielded a d of 1.11. The meta-analytic results reported in the Supplemental Material did not change regardless of the effect size used.
2. At Miami University, we believe that conditions for the RP:P protocol were assigned (by mistake) in a purely alternating fashion (i.e., anti-free-will, control, anti-free-will, control, etc.) rather than in a truly randomized order. The manipulation itself (for this protocol) was administered in a paper packet. Though the Miami team intended to shuffle the packets randomly before distributing them to participants, it appears that the packets instead were merely alternated, as the conditions in this protocol followed a perfectly alternating pattern in the data set. Results did not change when we excluded these data from the complete data set (see the Supplemental Material for analyses excluding these data).
3. Florian Brühlmann rebuilt the task to make it usable for modern computers.
4. We first had a single coder at each site assess suspicion, but during the post-data-collection review process, we had two coders recode the answers at each site. We report the results

of both sets of exclusion coding in the Supplemental Material. 5. The d values were calculated using Westfall, Kenny, and Judd's (2014) formula; the 95% confidence intervals are based on 1,000 bootstrap samples.

References

- Alquist, J., Ainsworth, S., & Baumeister, R. (2013). Determined to conform: Disbelief in free will encourages conformity. *Journal of Experimental Social Psychology*, 49, 80–86.
- Aristotle. (1980). *Nicomachean ethics* (W. D. Ross, Trans.). New York, NY: Oxford University Press.
- Aspinwall, L. G., Brown, T. R., & Tabery, J. (2012). The double-edged sword: Does biomechanism increase or decrease judges' sentencing of psychopaths? *Science*, 337, 846–849.
- Bakewell, S. (2017). *At the existentialist café: Freedom, being, and apricot cocktails with Jean-Paul Sartre, Simone de Beauvoir, Albert Camus, Martin Heidegger, Maurice Merleau-Ponty and others*. New York, NY: Other Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). doi:10.18637/jss.v067.i01
- Bell, B. A., Morgan, G. B., Schoeneberger, J. A., Loudermilk, B. L., Kromrey, J. D., & Ferron, J. M. (2010). *Dancing the sample size limbo with mixed models: How low can you go?* Retrieved from <https://support.sas.com/resources/papers/proceedings10/197-2010.pdf>
- Bird, A. (2007). Perceptions of epigenetics. *Nature*, 447, 396–398.
- Caspi, A., McClay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., . . . Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, 297, 851–854.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46.
- Crick, F. (1994). *The astonishing hypothesis*. New York, NY: Touchstone.
- Embley, J., Johnson, L. G., & Giner-Sorolla, R. (2015). *Reproducibility Project: Replication report – Replication of Study 1 by Vohs & Schooler (2008)*. Retrieved from <https://osf.io/uwt5f/>
- Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement*, 11, 93–103.
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., . . . Yamaguchi, S. (2011). Differences between tight and loose cultures: A 33-nation study. *Science*, 332, 1100–1104.
- Grant, M. J., Button, C. M., & Snook, B. (2017). An evaluation of interrater reliability measures on binary tasks using d-prime. *Applied Psychological Measurement*, 41, 264–276.
- Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution*, 5, 944–946.
- Kant, I. (1998). *Groundwork of the metaphysics of morals*. (M. Gregor, Ed. & Trans.). New York, NY: Cambridge University Press. (Original work published 1785)
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13). doi:10.18637/jss.v082.i13
- Martin, N. D., Rigoni, D., & Vohs, K. D. (2017). Free will beliefs predict attitudes toward unethical behavior and criminal punishment. *Proceedings of the National Academy of Sciences, USA*, 114, 7325–7330.
- Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Noûs*, 41, 663–685.
- Oishi, S. (2014). Socioecological psychology. *Annual Review of Psychology*, 65, 581–609.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349, Article aac4716. doi:10.1126/science.aac4716
- Paulhus, D. L., & Carey, J. M. (2011a). The FAD-Plus: Measuring lay beliefs regarding free will and related concepts. *Journal of Personality Assessment*, 93, 96–104.
- Paulhus, D. L., & Carey, J. M. (2011b). *A history of the FAD-Plus items*. Retrieved from http://www2.psych.ubc.ca/~dpaulhus/FAD_info/scale.development.details.doc
- Paulhus, D. L., & Margesson, A. (1994). *Free Will and Determinism (FAD) scale*. Unpublished manuscript, Department of Psychology, University of British Columbia, Vancouver, Canada.
- Protzko, J., Ouimette, B., & Schooler, J. (2016). Believing there is no free will corrupts intuitive cooperation. *Cognition*, 151, 6–9.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.
- Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., . . . Vohs, K. D. (2014). Free will and punishment: A mechanistic view of human nature reduces retribution. *Psychological Science*, 25, 1563–1570.
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). mediation: R package for causal mediation analysis. *Journal of Statistical Software*, 59(5). doi:10.18637/jss.v059.i05
- Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, 19, 49–54.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–1070.
- Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, 143, 2020–2045.