

5

Measuring and Manipulating Beliefs and Behaviors Associated with Free Will

THE GOOD, THE BAD, AND THE UGLY

*Jonathan Schooler, Thomas Nadelhoffer, Eddy Nahmias,
and Kathleen D. Vohs*

1. Introduction

Free will is typically considered a cornerstone of moral and legal responsibility. So what would happen if our belief in free will turned out to be nothing but a sophisticated illusion? One concern is that if such skepticism became widespread, the very pillars of morality could be shaken at their metaphysical foundations. This is not merely a tempest in an academic teapot. Recent findings from social psychology suggest that people who are exposed to anti-free will primes are more likely to cheat (Vohs & Schooler 2008) and behave aggressively (Baumeister, Masicampo, & DeWall 2009) than participants exposed to pro-free will or neutral primes (see also Baumeister et al. 2009; Rigoni, Soon, Sartori, & Brass 2011; Shariff et al. under review). In short, the gathering data suggest that people's believing they have free will could be instrumentally valuable from the dual standpoints of positive psychology and public morality, regardless of whether they actually have free will.

Since people's traditional views about agency and responsibility are increasingly challenged by some of the world's leading scientists and philosophers, we need to understand what effects (if any) these claims may have on people's beliefs and behaviors. Toward this end, our project set out to accomplish three primary objectives: (1) to develop a new psychometric instrument for measuring people's

beliefs specifically about free will, determinism, dualism, reductionism, responsibility, and punishment; (2) to generate new and more fine-grained tools for manipulating people's beliefs about free will and related concepts; and (3) to use these new tools to further explore the relationship between free will beliefs and both moral and nonmoral behavior. In this chapter, we describe our attempts to accomplish each of these goals. Because ours is still a work in progress, we do not draw any firm conclusions. While we have made some advances along the way, we have also run into some roadblocks and an occasional dead end. So our main goal is to candidly preview what we've found thus far—the good, the bad, and the ugly—in the hope that our work will motivate others to join us in our efforts to better understand the complex and fascinating web of beliefs and behaviors associated with free will and related concepts.

2. Measuring Beliefs about Free Will

It may turn out that there is no such thing as a consistent, unified folk concept of free will. There may just be a constellation of sometimes harmonious and sometimes conflicting beliefs about agency and responsibility. Given this possibility, we need psychometric tools that enable us to validly and reliably measure a wide range of beliefs about free will and related concepts. Unfortunately, the existing instruments are not well suited for our particular purposes.¹ So we decided to develop a new instrument for measuring free will beliefs.

The first step we took in constructing our scale was to come up with an extensive list of items that had better face validity than most items used in existing scales and that we predicted a priori would load onto common psychological factors (i.e., people's responses to these items would be statistically related in a way that suggests they measure an underlying belief or concept). The topics of the statements included in this initial list ranged widely from free will, responsibility, and punishment to the mind-body relationship, (in)determinism, fate, science, and the immaterial soul. We used exploratory factor analysis (EFA) to narrow down and fine-tune our items across four rounds of studies, with each round including at least 300 participants recruited through the paid panelist service offered by Qualtrics, which matches the sample to the demographics of the United States population. Between each round of data collection, we used EFA both to revise the items within the subscales and to eliminate some subscales altogether. Finally, after data collection involving more than 1,500 participants, we developed a 15-item scale, with three 5-item subscales for measuring the strength of people's beliefs about free will (FW), determinism (DE), and dualism/anti-reductionism (DU)² (see

Appendix for these items which comprise Part 1 of the Free Will Inventory; the FWI also includes 14 other items (Part 2) measuring beliefs about more complex relationships among free will, responsibility, choice, punishment, and predictability).

We used confirmatory factor analysis (CFA) to formally validate Part 1 of the FWI—see Nadelhoffer et al. (2014) for complete details. Here, we simply note that the CFA picked out three distinct factors that corresponded with our intended subscales, and the items that constitute each of the respective subscales loaded strongly together internally. The subscales themselves were not correlated, challenging earlier scales that presented belief in free will as contrasting with belief in determinism (e.g., Viney et al. 1982; Rakos et al. 2008). Instead, our findings support earlier work by Paulhus and Carey (2011) on the Free Will and Determinism (FAD+) scale concerning the independence of the psychological factors that appear to underlie people's beliefs about free will and determinism. On the surface, this provides some evidence for folk compatibilism (e.g., Nahmias et al. 2006).

On the other hand, we also found evidence suggesting that people ordinarily think having free will requires contra-causal agency (e.g., the unconditional ability to do otherwise inconsistent with determinism). For instance, in one sample ($N = 330$), participants responded to this statement in Part 2 of the FWI: "Free will is the ability to make different choices even if everything leading up to one's choice (e.g., the past, the situation, and their desires, beliefs, etc.) was exactly the same." We found that 79% agreed, suggesting that the case for folk compatibilism is on shakier ground than some have assumed. Furthermore, responses to this question correlated with scores on the free will and dualism subscales, and unlike scores on the free will subscale, responses were negatively correlated with scores on the determinism subscale—which suggests that this statement may be useful for distinguishing libertarians from compatibilists. At this point, given the conflicting data, advocates of the view that most people have beliefs that are most consistent with either incompatibilism or compatibilism have the burden of explaining away the conflicting data.

Given how intractable the traditional free will debate has been, the complicated empirical data on folk beliefs about free will and responsibility shouldn't come as a surprise. If we are to make further progress understanding people's beliefs about free will and related concepts, it is clear that we need to ensure that the psychometric tools we use for measuring these beliefs are up to the task. It is also clear that we need methods for manipulating free will beliefs that will enable us to explore the associated behaviors that interest us. We turn now to our efforts to develop such methods.

3. Manipulating Free Will Beliefs

In 2008, Vohs and Schooler reported a pair of studies demonstrating that manipulations that reduce people's belief in free will also increase the likelihood that they will display cheating behavior. Since this original demonstration, a number of additional studies have documented other negative consequences of anti-free will manipulations (Baumeister et al. 2009; Baumeister et al., 2009; Rigoni et al. 2011). Almost all of these studies have used the two different belief primes used by Vohs and Schooler (2008): (1) a one-page excerpt from Francis Crick's *The Astonishing Hypothesis* in which he argues that free will is an illusion (controls read a Crick excerpt that does not mention free will) or (2) the Velten procedure (1968) of having participants read and consider a series of 15 statements that are anti-free will, pro-free will, or neutral in content.

One feature of anti-free will primes used in prior studies is that they present a wide-range of potential challenges to free will, including various descriptions of determinism, reductionism, mechanism, and predictability, and various suggestions of fatalism and the irrelevance of conscious experiences. One of our goals was to try to isolate some of these features to discern which ones are most relevant to altering beliefs about free will. Ideally, experimentally testing among these features would involve designing primes that picked out and described specific features but not others and then measuring their impact on beliefs about free will (e.g., using the FWI). But given the stability of people's free will beliefs, we discovered that it may take powerfully worded primes that contain several threats to free will to alter people's beliefs and behaviors significantly. Nonetheless, we continue to test primes that highlight specific features (e.g., universal determinism, properly understood as a discovery about the laws of physics). Our hope is that these follow-up studies will help shed further light on which potential challenges to free will people find the most threatening and why. As things stand, while we know that people's beliefs and behaviors are sensitive to some threats to free will (and not others), we still don't have a very good understanding of the nature of the interaction.

Another one of our goals was to create 'matched pairs' of primes in the form of media articles that presented the same scientific information—for example, a functional magnetic resonance imaging (fMRI) study about brain activity during a complex decision-making task—and only varied regarding the scientists' claims about the implications of such studies for the existence of free will, as well as the respective headlines of the articles: "Neuroscientists Discover Free Will Is an Illusion" versus "Neuroscientists Discover How Free

Will Works.” In a pilot study using an earlier version of FWI, this anti-free will article lowered the mean scores on the FW subscale relative to the pro-free will article, even though the articles present the same scientific research (scores on the FAD+ did not vary). However, later studies using these news articles have typically not shown statistically significant effects on any of the subscales of the FWI and rarely had a significant impact on behavioral measures (discussed later). We also used this article format to create a no-soul prime (“Neuroscientists Discover the Soul is an Illusion”), in which the scientists interpreted the neuroscientific research to show that humans do not have nonphysical souls. This prime did not alter the scores on the DU or FW subscales of the early version of the FWI, though they raised the scores on the DE subscale.

So, while we continue to develop and test primes that present various other combinations of information challenging free will, our initial results suggest that altering beliefs about free will (as well as dualism and determinism) is more difficult than we had anticipated. When we limit the number of features presented as challenges to free will, or limit the force of the challenge, or alter some other aspect of the presentation (e.g., removing the Nobel-winning authority of Crick or removing the repetitious nature of reading 15 Velten statements), it is harder to diminish the strength of people’s reported belief in free will. To address this problem, we developed another prime designed to increase the impact of the anti-free will message.

This prime was based on an actual article published by the evolutionary theorist Jerry Coyne in *USA Today* (2012). Like the Crick essay, Coyne’s essay includes elements of determinism and reductionism (we are “collections of molecules that must obey the laws of physics”), and it suggests that we have no real choices (“You had no choice about making them [resolutions] and you’ll have no choice about whether you keep them”). As we will see, the Coyne prime often (though not always) altered scores on the final version of the FWI’s free will subscale and sometimes correlated with behavioral changes. For instance, in one online study, we asked 223 Qualtrics general-population participants to read one of three primes (the anti-free will article, pro-free will article, or Coyne article) and to offer a brief summary of the article in their own words. Participants then indicated their responses on a 100-point slider to a single question: “I have free will” and responded to the FWI. The results indicated that responses to the 100-point slider question differed significantly by prime (a mean of 84 with the pro-free will prime, a mean of 74 with the anti-free will prime, and a mean of 63 with the Coyne prime). Participants reading the Coyne article had lower scores on the FW subscale

than those reading the pro-free will article and than those reading the anti-free will article. Scores on the DE and the DU subscales did not differ significantly among the primes, perhaps suggesting that at least the Coyne prime diminishes belief in free will without doing so by influencing beliefs regarding dualism or determinism.

Our initial attempts to develop methods for measuring beliefs about free will and for manipulating those beliefs suggest that beliefs in free will are typically robust and difficult to alter, and that these beliefs, including their relation to potential challenges, are highly complex. Depending on these relations among beliefs, various challenges to free will are likely to have varying impacts on different people. While we continue to refine our methods of sorting out some of the relations among beliefs about free will and these other concepts and of testing what information challenges those beliefs, our tentative conclusions are that the challenges must be presented as substantial and multifaceted, and this is suggested as well by the behavioral studies, to which we now turn.⁴

4. Manipulating Behavior

The behavioral studies we ran were motivated by the goal of distinguishing between two alternative mechanisms that may contribute to the impact of anti-free will messages on behavior. According to an exoneration account, telling people they lack free will provides them an excuse for acting badly. According to the ego-depletion account, discounting free will or agency deflates people's sense of personal free will, thereby undermining their ability or motivation to exert self-control. To assess these accounts we examined the degree to which ego depletion alone is sufficient to characterize the evidence. If experimental findings point to a role for ego depletion in mediating the behavioral impact of anti-free will primes, then parsimony favors this account over one that alternatively (or additionally) posits exoneration. In contrast, if the findings fail to fully support an ego-depletion finding, then this provides indirect evidence for alternative views such as the exoneration account.

If we assume for the sake of argument that the ego-depletion account is true, we can formulate the following testable hypotheses and predictions:

Hypothesis 1 (H1): If discouraging belief in free will is ego depleting, then there should be parallels between the effects of anti-free will

messages and ego-depletion manipulations. In other words, anti-free will messages should impair performance on the same types of tasks that have previously been found to be disrupted by ego depletion.

Hypothesis 2 (H₂): If discouraging a belief in free will undermines people's sense of personal free will, then following such manipulations, individuals may report a reduced sense of agential control. In other words, anti-free will manipulations may reduce people's willingness to positively endorse items on scales such as the Self-control scale (Tangay, Baumeister, and Boone, 2004) or the Locus of control scale (Rotter, 1966).

Hypothesis 3 (H₃): If there are strong parallels between ego depletion and challenging beliefs in free will, then ego depletion may reduce people's free will beliefs in a manner similar to that observed with anti-free will messages.

Two different strands of behavioral studies were conducted to address these hypotheses: one by Kathleen Vohs and her colleagues at the University of Minnesota (UMN) and a second by Jonathan Schooler and his colleagues at the University of California Santa Barbara (UCSB). Both strands were motivated by the goal of assessing the source of the impact of anti-free will messages on behavior as outlined earlier. While the results provide some support for all three hypotheses, several of the studies failed to find any behavioral effects of the anti-free will messages, and the impact of these messages on people's beliefs was somewhat inconsistent. A number of possible reasons could explain the various unpredicted null findings and the discrepancies between studies that we consider in telling this part of our interdisciplinary tale.

4.1. Strand One: University of Minnesota (UMN) Behavioral Studies

The UMN strand involved four studies that explored H₁, namely, that anti-free will messages should impair performance on the same types of tasks previously found to be disrupted by ego depletion. Consequently, we examined the impact of anti-free will messages on three types of tasks that involve self-control: controlling unwanted thoughts, math problem solving, and helping behavior (Baumeister et al., 2009). Finally, an additional experiment examined a performance domain—creativity—that has not been associated with self-regulation and for which there is some evidence that a deficit in self-regulation might actually be beneficial.

4.1.1. UMN Experiment 1: Thought Suppression

If anti-free will primes affect behavior in a manner similar to the effects of ego depletion, it follows that they should also undermine performance on measures of mental control. One classic index of mental control is attempting to keep an unwanted thought out of consciousness (Wegner, 1989). Indeed, thought suppression is regularly used to induce ego depletion (Burkley, 2008; Muraven et al., 1998). Accordingly, we reasoned that if anti-free will messages lead to ego depletion, participants who receive anti-free will messages should be less successful at avoiding unwanted thoughts relative to those who are given pro-free will messages.

Participants ($N = 63$) came to the laboratory and were told the session included two different studies, one on reading comprehension and one marketing study. This two-experiment procedure is common and is used to reduce the chances that participants will connect the first and second parts of the study. But it is not the method used in all experiments manipulating belief in free will, including some we report later, and this variation may make a difference.

Participants read either the pro- or anti-free will news article described earlier or a control essay, and reported their beliefs on an earlier version of the FWI. Then, believing they had moved on to the second study, participants were given a task involving thought suppression, specifically, trying not think about a purple elephant. The key dependent measures were scores on the FWI and number of intrusions of the unwanted thought.

The anti-free will messages marginally reduced people's belief in free will, but contrary to the predictions of an ego-depletion account, there was no difference between the anti-free will and pro-free will messages on participants' ability to keep unwanted thoughts at bay. As always, there are many possible explanations for null findings. First, it may be that anti-free will messages simply do not undermine mental control in the manner that would be expected if they cause ego depletion. Alternatively, it could be that for some reason unwanted thought intrusion is an insensitive measure of ego depletion. Finally, the anti-free will prime may not have induced changes in beliefs sufficiently to produce behavioral differences, at least with the statistical power afforded by the relatively low number of participants per condition in this study.

4.1.2. UMN Experiment 2: Math Problem Solving

A number of studies have found that ego-depleting tasks can disrupt participants' performance on math problems (Hagger et al., 2010). The basic idea

is that math problems require mental control both for a person to continue performing an unenjoyable task and also to avoid careless errors. If anti-free will messages undermine mental control in a manner similar to ego depletion, then it would be expected that they would similarly disrupt performance on math problems. UMN Experiment 2 examined this issue.

After reading the same cover story about doing two experiments and reading one of the two news articles or a control essay, participants ($N = 60$) completed an earlier version of the FWI and then were given several multi-digit addition and subtraction problems. Participants were told that they could work on them to exercise their brains but could quit at any point.

Participants in the anti-FW condition reported marginally lower scores on the free will subscale, and significantly higher determinism scores than participants in the pro-free will condition. As in previous studies there was no effect on the dualism subscale. The results provide some evidence in support of a relationship between anti-free will messages and math performance. Participants who received the anti-free will message performed worse on math problems than those who received the pro-free will message—worse not in terms of persistence but rather of efficiency, in that the pro-free will group achieved a similar number of correct answers faster than the anti-free will group. While this finding does not fully line up with an ego-depletion account, proponents of such a view would presumably predict that depleted participants would put less effort into the process, which is a common finding in the depletion literature (Baumeister, Vohs, & Tice, 2007). So it is possible that this account might still help explain the findings.

4.1.3. UMN Experiment 3: Helpfulness

Helpfulness is another self-regulatory behavior associated with ego depletion (Baumeister, Vohs, & Tice, 2007). Indeed, one previously published study found that both manipulated and measured free will beliefs were negatively related to helpfulness (Baumeister et al., 2009). Thus, there were good conceptual and empirical reasons to expect that an anti-free will manipulation would reduce helpfulness compared to other manipulations.

Participants ($N = 45$) read either the pro-free will or anti-free will news article or control article, and then completed a version of the FWI. The helping opportunity came via a request from a confederate who claimed confusion and asked for assistance (Vohs, Mead, & Goode, 2006). How long participants helped the confederate (surreptitiously timed) served as the measure of helpfulness.

Although participants who received the anti-free will messages reported a reduced belief in free will, there was no effect of this manipulation on helpfulness. As before, an inability to reject the null hypothesis could stem from many factors. One possibility stems from the issue of statistical power. Another possibility has to do with the manner in which the anti-free will messages were introduced. It is notable that Baumeister et al. (2009) found a negative effect of anti-free will messages on helping behavior using the Velten procedure rather than the news article procedure used here, and the Velten procedure may influence beliefs or behavior more than the news article. This was one of the issues we tried to address with the following study.

4.1.4. UMN Experiment 4: Cheating and Creativity

This experiment had several goals. First, given the equivocal impact of the anti-free will news articles in the prior studies, we aimed to compare the effect of that manipulation to the effect of the Velten procedure, which has been shown to impact behavior in a number of published studies (Alquist et al. 2013; Baumeister et al. 2009; Vohs & Schooler, 2008). In order to maximize the likelihood of seeing effects of our manipulations, we selected a dependent measure (cheating) that has previously been sensitive to primes. Second, we were interested in examining the impact of anti-free will primes in a domain for which an ego-depletion account would seem less likely to produce effects—namely, creativity. Unlike many other areas of intellectual accomplishment, creativity does not necessarily benefit from exercising executive resources and self-regulation. For example, manipulations that dampen self-regulatory abilities, such as giving participants alcohol, can enhance creative problem solving (e.g. Jarosz, Colflesh, & Wiley, 2012). Thus, if the primary effect of anti-free will messages is to dampen self-regulatory processes, then they should not impair and might even enhance creative processing.

Participants ($N = 59$) completed one of the free will belief manipulations—namely, the pro-free will or anti-free will news articles or the pro-free will or anti-free will Velten statements, or read a control essay. Participants then completed a version of the FWI. Finally, believing they were engaged in a separate study, participants completed three measures: a die rolling task intended to assess cheating (modeled after Greene & Paxton 2009), and both a convergent and a divergent creativity task. In the convergent task, participants performed the Remote Associates Test (Mednick, 1968), which requires participants to recognize the common associate of three otherwise unrelated words. The divergent task was the Draw-An-Animal task, whereby participants are asked to draw an animal from a planet that is very different from Earth.

The results once again provided a mixed picture of the impact of primes on behavior and beliefs. With respect to cheating, we found evidence of a modest effect as the pro-free will Velten statements marginally reduced cheating compared to the effect of the other conditions, consistent with the possibility that the Velten manipulation has greater impact than others. At the same time, although the difference between the Velten conditions is a conceptual replication of the earlier Vohs and Schooler findings, it is only partial since in this study the result was driven by the pro-free will Velten statements, whereas previously the anti-free will primes drove the effect. It is difficult to know precisely why this disparity occurred but we note that Trager, Vallacher, & Sherman (2013) recently reported two replications of manipulated free will beliefs influencing cheating—and measured cheating at the individual level. Thus, the general hypothesis that a relative difference in free will beliefs affects cheating seems supported by the existing data.

With respect to creativity, although there were no overall differences between conditions, participants in the anti-free will Velten condition were marginally less creative on both measures relative to the other conditions, suggesting that future research on this topic would be fruitful.

4.2. Strand Two: University of California at Santa Barbara (UCSB) Behavioral Studies

There were several important similarities and differences between the UMN and UCSB strands of studies. Like the UMN strand, the UCSB strand (1) was aimed at investigating how anti-free will primes affect behavior and beliefs, (2) explored the hypothesis (H₁) that anti-free will primes disrupt domains previously associated with ego depletion, and (3) sought to determine which types of primes are particularly likely to affect behavior and/or beliefs. However, unlike the UMN strand, the UCSB strand (1) conducted the studies using a web-based mTurk platform that forgoes some laboratory control in favor of increased sample size, (2) investigated two additional implications (H₂ and H₃) of the conjectured interactions between anti-free will messages and ego depletion, and (3) explored the possible consequences of varying whether the prime and dependent behavioral measures are characterized as being part of the same experiment or as two different experiments.

4.2.1. UCSB Experiment 1: Cheating and Self-Control

The goals of this experiment were twofold. First, we wanted to use a web-based experimental paradigm that affords large numbers of participants to conceptually

replicate Vohs and Schooler's (2008) finding that anti-free will sentiments increase cheating. One of the possible limitations to prior studies examining the impact of anti-free will primes is statistical power. Thus, there are real advantages to identifying behavioral markers of cheating that can be assessed using the large samples available with Web-based platforms. In addition to examining the impact of anti-free will manipulations on cheating, we were also interested in its impact on people's sense of personal or agential control. As H₂ states, if anti-free will messages increase anti-social behavior by reducing ego resources, then they might also reduce people's sense of control as measured by the Self-control scale (Tangney, Baumeister & Boone, 2004), the Locus of control scale (Rotter, 1966), and the mind-wandering questionnaire (Mrazek et al., 2013). Moreover, even if there is no impact of the primes on people's endorsement of various measures of self-reported agency, it is useful to examine the relationship between those measures and belief in free will. Accordingly, if ego depletion and belief in free will are related, then we might expect a relationship between people's belief in free will and their perceived sense of agency.

Participants included 184 individuals living in the United States recruited to participate in an online experiment through mTurk. Participants first read either the anti-free will article by Coyne described earlier or a neutral essay. Next, participants either worked on an anagram task or completed various questionnaires (counterbalanced across subjects). Participants were explicitly told to work on the anagram task for 10 minutes. A timer was present indicating how much time had elapsed. However, a "next" button enabled participants to cheat by terminating the program early. The questionnaires included a single-item belief in free will question (the 100-point free will slider described earlier), the three subscales from Part 1 of the FWI, the Self-control scale, and the Locus of control scale.

Relative to participants who read a neutral article, participants exposed to the anti-free will article spent significantly less than the required amount of time working on the task. In addition, anti-free will participants showed a reduced degree of endorsement of "I have free will" on the free will slider relative to control participants. However, there was no effect of the primes on any of the other free will measures or the various measures of self-control. Finally, scores on the Self-control scale, intrinsic locus of control, and mind-wandering questionnaire were correlated with the scores on the free will subscale of the FWI.

4.2.2. UCSB Experiment 2: Anagram Completion and Ego Depletion

This experiment sought to compare the effect of ego depletion and anti-free will primes on a measure (anagram completion) known to be susceptible

to ego depletion. If undermining belief in free will impacts performance by increasing ego depletion, then its effect should be similar to that of an actual ego-depletion task (Schmeichel, 2007). In this case, we compared the effect of having participants read either anti-free will or neutral primes to the effect of having them write an essay that either required the avoidance of the letters “a” and “n” (a standard ego-depleting task) or did not require that. In addition, we also explored the further issue (H₃) of whether ego depletion might reduce people’s belief in free will.

Participants were 800 US participants from mTurk. The design included two free will conditions (anti-free will Coyne essay or neutral essay) and two ego-depletion conditions (writing a short story without using the letters a and n or with no constraints). After reading the essays or writing their story, participants either worked on an anagram task or completed the various free will and self-control questionnaires and then completed the other activity (counterbalanced). The anagram task differed from that used in the previous study in that participants were not given the option to prematurely terminate the task.

The results replicated the previously observed effect of ego depletion on anagrams (Schmeichel, 2007) such that participants completed fewer anagrams following the challenging essay task relative to the unconstrained task. Receiving the anti-free will message also marginally reduced anagram performance relative to the control condition. Participants in the ego-depletion condition also indicated a significantly lower belief in free will on the slider relative to controls. As in UCSB Experiment 1, there was an effect of the anti-free will prime on the free will slider but not on any of the other subscales of the FWI, nor on the self-control measures. Once again there was a relationship between belief in free will and the self-control measures, such that the more people believed in free will, the more they tended to report possessing more self-control and intrinsic locus of control and less mind wandering.

4.2.3. UCSB Experiment 3: The Compromise Effect

The goal of this study was to explore the possible impact of anti-free will messages on the compromise effect—that is, the tendency for people to prefer an option when it is presented as a middle option (such as option b in a set of options a-b-c) rather than when it is an extreme option (such as option b in a set of options b-c-d) (Simonson, 1989). Research has found that when individuals engage in extensive deliberation they are more susceptible to the compromise effect than when they engage in more superficial heuristic processing. We therefore reasoned that if anti-free will messages undermine individuals’

capacity for (or motivation to expend) resource-demanding deliberative processes, as the ego-depletion account would suggest, that may reduce the potential impact of the compromise effect.

In addition to exploring the compromise effect in this study, we also sought to consider two factors that might have contributed to the generally smaller effects in the aforementioned UMN strand of studies relative to the earlier work by Vohs, Schooler, Baumeister, and others. First, because Vohs and Schooler (2008) used an anti-free will essay written by Francis Crick (1994), it seemed possible that this essay might be more effective than the Coyne essay and news story used in previous studies. To address this possibility, UCSB Experiment 3 compared the effects of both the Crick and the Coyne essay to a neutral control. Second, a discrepancy was identified in the manner that the prime was introduced in the UCSB and the UMN strands. Specifically, the UMN-based studies introduced the anti-free will messages as a completely different experiment from the various behavioral measures. Although not mentioned in the original Vohs and Schooler (2008) manuscript, this division was also used in those studies. Previous studies from a number of domains, including mere exposure (Bornstein & Agostino, 1994), terror management (Greenberg, Pyszczynski, et al. 1994), and self-affirmation (Sherman, Cohen, et al., 2009), report that primes are more effective when they are introduced in a context distinct from the one in which the priming effects are expected to occur. It thus seemed possible that this difference in paradigms might explain why the UMN studies found some effects of the primes on various measures of belief in free will, even with relatively small sample sizes, while the UCSB studies did not. To explore this possibility, UCSB Experiment 3 characterized the essay reading task and the decision making/scale completion components of this study as involving two entirely different experiments.

Participants included 314 individuals from the United States recruited through mTurk. Participants were told that they would be engaging in two short separate studies. The introduction of the primes was framed as part of the first study in which participants were asked to answer a variety of mood questions, read and summarize an article (one of the two anti-free will essays or a neutral essay), and then answer additional questions about their mood. Once they had completed this task they were thanked, informed study one was complete, and asked to move on to the next page for the instructions to the second study.

The behavioral measures were framed as a second study involving consumer decision making about laptops. Participants were introduced to three brands of laptops labeled “Brand A,” “Brand B,” or “Brand C,” each to

be chosen from a group of four laptops. Condition 1 had laptops 1, 2, and 3 labeled A, B, and C, respectively. Condition 2 had laptops 2, 3, and 4 labeled A, B, and C, respectively. With regard to the compromise effect it would be expected that in Condition 1 participants would be more likely to prefer laptop 2 over laptop 1 or 3 because it was the middle option. In Condition 2 it would be expected that participants would prefer their middle option, laptop 3. Following subjects' consumer decisions, they were asked to complete a series of scales including the free will slider, the FWI, and several self-control scales.

There were a number of informative findings. First, there was no evidence that either of the anti-free will messages reduced the compromise effect. Given that the compromise effect is known to rely on the engagement of deliberative processes (Simonsohn, 1989), and given that ego depletion is known to dampen such processes, this experiment failed to provide further support for the ego-depletion account. At the same time, to our knowledge no study has specifically examined the effect of ego depletion on the compromise effect, so even though it should in principle be susceptible to ego depletion, it is possible that it simply is not. Another possibility is that the effects of anti-free will messages, though resembling ego depletion in some respects (as suggested by previous studies), differ from ego-depletion in other respects. Clearly, further research is needed on this issue.

A second informative result was the impact of the anti-free will primes on responses to both the FWI and the various measures of agential control. In general, the Crick prime had a somewhat greater impact on participants' responses to the free will and agential control measures. Relative to both the neutral and the Coyne primes, the Crick prime significantly increased individuals' determinism scores and reduced scores on the self-control and the internal locus of control measures. This finding suggests that part of the reason our various labs found attenuated effects of the primes is that the new primes pose less substantial challenges to free will than the primes used in the original Vohs and Schooler study.

Given the disparity between these studies, it is possible that a change in the presentation of the primes—that is, as part of a separate study—may also have been responsible for their greater impact on reported beliefs. Specifically, when the free will primes are presented in too obvious proximity to the behavioral and questionnaire measures, it may be that reactance takes place and people resist adopting the attitudes presented by the primes. In contrast, by separating the primes, people may (at least temporarily) incorporate the anti-free will messages into their implicit worldview, which may thereby enhance the impact of the primes.

4.2.4. UCSB Experiment 4: Cheating and Framing Effects

This experiment was designed to test whether framing the primes as part of a separate study can increase their efficacy. In this study, the prime and behavioral aspects of the study were framed as a single experiment for half of the participants (as in UCSB Experiments 1 and 2) and as two separate experiments for the remaining participants (as in UCSB Experiment 3 and the UMN studies). We hypothesized, based on the different outcomes of the studies reported here, that the primes would have a larger effect when they were presented as a separate experiment from the other measures compared to when they were characterized as part of the same experiment.

Participants ($N = 414$) were either told that they were participating in a single study or that they were going to work on two different studies, one involving the relationship between reading and mood, and another looking at the relationship between people's beliefs and their performance on various tasks. Participants then received either the anti-free will (Crick) or neutral prime, followed by a popular mood scale (Watson, Clark, & Tellegen, 1988). Next, participants were given the various free will and self-control scales and the cheating paradigm (counterbalanced), adopted from Greene and Paxton (2009). This task involved asking participants to guess the outcome of coin flips, and then, after being told the outcome, report whether their guess was correct. To encourage cheating, participants were told they would receive a monetary reward for every correct answer.⁵

The results of UCSB Experiment 4 indicate that whether the primes and measures are characterized as one study or two separate studies can significantly modulate the impact of the primes. On several of the key measures, including belief in free will, and external locus of control, the effects of the anti-free will prime were only observed when the primes and the measures were framed as distinct experiments. Although there was a marginal trend for a cheating effect in the two-experiment condition, the present study again failed to find a significant difference between the anti-free will and neutral primes. It is difficult to know why UCSB Experiment 4 failed to fully conceptually replicate the impact of anti-free will essays on cheating. One possibility is that the coin flip measure of cheating was quite different from either of those used by Vohs and Schooler. It is also possible that participants on mTurk believed or learned that they would get the full amount regardless of their performance (via online Web forums), and this may have undermined their motivation to cheat. Even though Experiment 4 was unsuccessful in eliciting an effect of priming on cheating, its successful demonstration of the framing effect regarding the impact of the primes provides a potentially important

step forward in our understanding of the conditions under which priming is most likely to occur. We hope that these insights will help us to develop more effective priming studies in the future.

5. Lessons Learned

Two kinds of lessons emerge from our bumpy journey toward understanding the measurement and manipulation of beliefs and behaviors associated with free will: what we learned about the topic itself and some more general lessons about the investigative process that contributed both to the bumps and possibly to their eventual explication. We review these two themes in turn.

With respect to the measurement of beliefs about free will and related concepts, we successfully developed a philosophically nuanced metric of beliefs in free will that disentangles three distinct constructs (free will, determinism, and dualism), and that revealed strong psychometric properties—part 1 of the FWI—and that also provides information about relationships among these beliefs and between them and related beliefs—part 2 of the FWI. We developed new primes in an attempt to manipulate beliefs about free will, and not surprisingly, we found that it is difficult to reduce belief in free will without strong and multifaceted messages about why free will is an illusion (as in the Crick prime and the Velten statements).

With respect to assessing the impact of free will primes on attitudes and behavior, we found three lines of evidence for the conjecture that anti-free will primes undermine agential control in a manner similar to that of ego depletion. Consistent with H₁, we found that anti-free will primes disrupt a number of domains that have also been vulnerable to ego depletion, including math performance (UMN Experiment 1) and anagram solutions (UCSB Experiment 2). Consistent with H₂, we found that belief in free will is associated with the perception of agential control as measured by relevant scales (UCSB Experiments 1, 2, and 3) and that anti-free will primes can actually reduce people's perception of their own agential control (UCSB Experiments 3 and 4). Finally, consistent with H₃, we found a symmetrical effect such that participating in an ego-depleting task can reduce people's belief in free will (UCSB Experiment 2).

Although these studies provide further support for the hypothesized relationship between anti-free will primes and ego depletion, questions remain about the nature of this relationship. For example, we failed to find effects of anti-free will primes on a variety of tasks that have been (or could reasonably be expected to be) associated with ego depletion in the past, including thought

intrusions (UMN Experiment 1), helping behavior (UMN Experiment 3), and the compromise effect (UCSB Experiment 3). Furthermore, we observed marginal negative effects of anti-free will primes on creativity (UMN Experiment 4), which is a domain that has not previously been associated with ego depletion, and indeed represents an area where dampening of self-regulatory abilities has been found to be helpful (Jarosz et al. 2012; White & Shaw, 2006, 2011). Given these equivocal findings it seems that we must conclude that the jury is still out on whether ego depletion provides an adequate account of the impact of anti-free will primes (when they are observed). Clearly more research is needed to further explore this relationship.

One of the inescapable conclusions from this series of studies is that the effects of free will primes on attitudes and behavior is more fickle and difficult to measure than we had originally anticipated. Our initial attempts at using more specific anti-free will primes to illustrate particular threats to free will typically failed to have a significant impact on reported beliefs, and the more strongly worded primes, while more effective, still failed to produce significant effects on some of the free will measures in a number of the studies. Moreover, while the anti-free will primes did affect behavior in a number of cases, in several studies we failed to replicate previously observed findings, including the negative effects of anti-free will essays on helping (UMN Experiment 2) and cheating (UMN Experiment 4, UCSB Experiment 4).

Given the challenges that we experienced in replicating some previously reported effects, some might be tempted to dismiss this line of research as too fickle to be meaningful. Indeed, this is an increasingly common response when researchers encounter replication difficulties (Pashler & Harris, 2012). However, the present series of studies suggests that a different conclusion is warranted. Although we failed to get significant effects in several cases, we also found significant effects in a number of other cases. Most important, we identified a host of factors that may contribute to the now-you-see-it-now-you-don't quality of some of the effects that we investigated. These include both the strength of the primes (Qualtrics studies with FWI, UMN Experiment 4, UCSB Experiment 3) and whether the primes and their subsequent measurements were characterized as part of a single experiment or separate experiments (UCSB Experiments 3 and 4).

The impact of contextualizing the primes and measures within a single or separate study highlights an important psychological source of the variability of experimental outcomes in this and other paradigms. Specifically, this finding is consistent with findings from various

literatures, including mere exposure (Bornstein & Agostino, 1994), terror management (Greenberg et al. 1994), and self-affirmation (Sherman et al., 2009), which demonstrate that psychological effects can be greater when participants are unaware of their source. It seems likely that a host of factors could affect whether participants' perceive a relationship between the primes and their subsequent responses. Such factors may not only be an important source of the variability in our ability to replicate the effect of anti-free will primes but they may also underpin many of the replication issues that have faced the priming literature of late (Doyen, Klein, Pichon, & Cleeremans, 2012; Pashler, Coburn, & Harris, 2012). Fortunately, these are empirically testable conjectures.

We hope that our forthright presentation of the results from this set of experiments will help researchers better measure beliefs about free will and further explore the possible interactions between beliefs about free will and various behaviors, as well as the underlying causes of those interactions.

Acknowledgments

The material presented in this chapter is based on work supported in part by a grant to the four authors from the Big Questions in Free Will project of the John Templeton Foundation, and in part by the John D. and Catherine T. MacArthur Foundation [TN], the Law and Neuroscience Project [TN], and the Regents of the University of California [JS]. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the John Templeton Foundation, the John D. and Catherine T. MacArthur Foundation, the Law and Neuroscience Project, or of the Regents of the University of California. We would also like to thank the following people for their contributions on the studies discussed in this chapter: Chandra Sripada, Jason Shepard, Lisa Ross, Brett Ouimette, Michael Mrazek, and Kelly Rowe. Finally, we would like to thank Al Mele for his wonderful leadership on the Big Questions in Free Will project.

APPENDIX: FREE WILL INVENTORY (FWI) PART I

The Free Will Subscale (FW):

1. People always have the ability to do otherwise.
2. People always have free will.
3. How people's lives unfold is completely up to them.
4. People ultimately have complete control over their decisions and their actions.

AQ: Please confirm the placement of Appendix is OK?

5. People have free will even when their choices are completely limited by external circumstances.

The Determinism Subscale (DE):

1. Everything that has ever happened had to happen precisely as it did, given what happened before.
2. Every event that has ever occurred, including human decisions and actions, was completely determined by prior events.
3. People's choices and actions must happen precisely the way they do because of the laws of nature and the way things were in the distant past.
4. A supercomputer that could know everything about the way the universe is now could know everything about the way the universe will be in the future.
5. Given the way things were at the Big Bang, there is only one way for everything to happen in the universe after that.

The Dualism/Anti-Reductionism Scale (DU):

1. The fact that we have souls that are distinct from our material bodies is what makes humans unique.
2. Human action can only be understood in terms of our souls and minds and not just in terms of our brains.
3. Each person has a non-physical essence that makes that person unique.
4. The human mind cannot simply be reduced to the brain.
5. The human mind is more than just a complicated biological machine.

NOTES

1. For a more thorough treatment of the limitations and shortcomings of the extant scales for measuring free will beliefs, see Nadelhoffer, Shepard, Nahmias, Sripada, & Ross (2014).
2. Anti-reductionism, as a thesis about the inability to explain mental states in terms of brain states, does not entail substance dualism. However, because responses from the general population loaded onto a common factor, our subscale includes items representing both concepts. For simplicity, we will refer to it as the dualism subscale.
3. The one exception is Shariff et al. (in preparation), which uses news articles.
4. The authors have differing views regarding what features of existing primes are most likely affecting people's beliefs and behaviors.
5. Due to restrictions by the Institutional Review Board, all participants actually received the maximum possible reward though they did not learn this until the completion of the experiment.

REFERENCES

- Alquist, J. L., Ainsworth, S. E., & Baumeister, R. F. (2013). Determined to conform: Disbelief in free will increases conformity. *Journal of Experimental Social Psychology*, 49, 80–86.
- Baumeister, R. F., Vohs, K. D. & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16, 351–355.
- Baumeister, R., Masicampo, E., and DeWall, C. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin*, 35(2), 260–268.
- Baumeister, R., Mele, A., & Vohs, K. (Eds.). (2010). *Free Will and Consciousness: How Might They Work?* Oxford: Oxford University Press.
- Bornstein, R., & D'Agostino, P. R. (1994). The attribution and discounting of perceptual fluency: Preliminary tests of a perceptual fluency/attributional model of the mere exposure effect. *Social Cognition*, 12(2), 103–128.
- Burkley, E. (2008). The role of self-control in resistance to persuasion. *Personality and Social Psychology Bulletin*, 34, 419–431.
- Crick F. (1994). *The Astonishing Hypothesis*. New York: Scribner's.
- Coyne, J. (2012). Why you don't really have free will. *USA Today* (1/1/2012).
- Doyen S., Klein O., Pichon C.-L., & Cleeremans A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE* 7 (1) e29081.
- Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L. & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts in mortality salience effects. *Journal of Personality and Social Psychology* 67(4), 627–637.
- Greene, J. D. & Paxton, J. M. (2009). Patterns of neural activity associated with honest and dishonest moral decisions. *Proceedings of the National Academy of Sciences*, 106, 12506–12511.
- Hagger, M. N. Wood, C., Stiff, C., & Chatzisarantis, N. L. D. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, 136(4), 495–525.
- Jarosz, A. F., Colflesh, G. J. H., & Wiley, J. (2012). Uncorking the muse: Alcohol intoxication facilitates creative problem solving. *Consciousness and Cognition*, 21(1), 487–493.
- Mednick, S. A. (1968). The remote associates test. *Journal of Creative Behavior*, 2, 213–214.
- Mrazek, M. D., Phillips, D. T., Franklin, M. S., Broadway, J. M., & Schooler, J. W. (2013). Young and restless: Validation of the Mind-Wandering Questionnaire (MWQ) reveals disruptive impact of mind-wandering for youth. *Frontiers in Psychology*, 4, 560.
- Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, 74(3), 774.

- Nadelhoffer, T., Shepard, J., Nahmias, E., Sripada, C., & Ross, L. (2014). The Free Will Inventory: Measuring Beliefs about Agency and Responsibility. *Consciousness and Cognition*, 25, 27–41.
- Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research*, 73, 28–53.
- Pashler, H., Coburn, N., & Harris, C. R. (2012). Priming of social distance? Failure to replicate effects on social and food judgments. *PLoS ONE*, 7(8): e42510.
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science*, 7(6), 531–536.
- Paulhus, D., & Carey, J. (2011). The FAD–Plus: Measuring lay beliefs regarding free will and related constructs. *Journal of Personality and Assessment*, 11(1), 96–104.
- Rakos, R., Laurene, K., Scala, S., & Slane, S. (2008). Belief in free will: Measurement and conceptualization innovations. *Behavior and Social Issues*, 17, 20–39.
- Rigoni, D., Wilquin, H., Brass, M., & Burle, B. (2013). When errors do not matter: Weakening belief in intentional control impairs cognitive reaction to errors. *Cognition*, 127, 264–269.
- Rigoni, D., Kuhn, S., Gaudino, G., Sartori, G., & Brass, M. (2012). Reducing self-control by weakening belief in free will. *Consciousness and Cognition*, 21, 1482–1490.
- Rigoni, D., Kuhn, S., Sartori, G., & Brass, M. (2011). Inducing disbelief in free will alters brain correlates of preconscious motor preparation: The brain minds whether we believe in free will or not. *Psychological Science*, 22, 613–618.
- Rotter, J. (1966). Generalized expectancies for internal versus external control of reinforcements. *Psychological Monographs*, 80, Whole No. 609.
- Schmeichel, B. J. (2007). Attention control, memory updating, and emotion regulation temporarily reduce the capacity for executive control. *Journal of Experimental Psychology: General*, 136, 241–255.
- Schooler, J. W. (2011). Unpublished results hide the decline effect. *Nature*, 470, 437.
- Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J., Clark, C., Schooler, J. W., Baumeister, R. F., & Vohs, K. D. (under review). Free will and punishment: A mechanistic view of human nature reduces retribution.
- Sherman, D. K., Cohen, G. L., Nelson, L. D., Nussbaum, A. D., Bunyan, D. P., & Garcia, J. (2009). Affirmed yet unaware: Exploring the role of awareness in the process of self-affirmation. *Journal of Personality and Social Psychology*, 97, 745–764.
- Simonson, I. (1989). Choice Based on Reasons: The Case of Attraction and Compromise Effects. *Journal of Consumer Research*, 16, 158–174.
- 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, 271–324.
- Trager, B. M., Vallacher, R. R., & Sherman, R. A. (2012). The consequences of disbelief in free will: Diminished morality or enhanced conformity. Poster presented at meeting of the Society for Personality and Social Psychology, New Orleans, LA.

- Velten, E. (1968). A laboratory task for the induction of mood states. *Behavioral Research and Therapy*, 6, 607–617.
- Viney, W., Waldman, D., & Barchilon, J. (1982). Attitudes towards punishment in relation to beliefs in free will and determinism. *Human Relations*, 35, 939–950.
- Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science*, 314, 1154–1156.
- Vohs, K. D. & Schooler, J. S. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, 19(1), 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.
- Wegner, D. M. (1989). *White bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control*. New York: Viking/Penguin.
- White, H. A., & Shah, P. (2006). Uninhibited imaginations: Creativity in adults with Attention-Deficit/Hyperactivity Disorder. *Personality and Individual Differences*, 40(6), 1121–1131.
- White, H. A., & Shah, P. (2011). Creative style and achievement in adults with attention-deficit/hyperactivity disorder. *Personality and Individual Differences*, 50(5), 6